Intelligent Pilot Aids for Flight Re-Planning in Emergencies

Final Report
NAG 1 01055
NASA Langley Research Center
Hampton VA 23681-0001

for the period

March 1, 2001 – April 25, 2005.

Attention: Anna C. Trujillo
757-864-8047

Amy R. Pritchett, Sci.D., Principal Investigator
Jennifer J. Ockerman, PhD, Co-Principal Investigator
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
(Tel) 404-894-0199
(Fax) 404-894-2301
Amy.Pritchett@isye.gatech.edu

August 2005

Abstract

Effective and safe control of an aircraft may be difficult or nearly impossible for a pilot following an unexpected system failure. Without prior training, the pilot must ascertain on the fly those changes in both manual control technique and procedures that will lead to a safe landing of the aircraft. Sophisticated techniques for determining the required control techniques are now available. Likewise, a body of literature on pilot decision making provides formalisms for examining how pilots approach discrete decisions framed as the selection between options. However, other aspects of behavior, such as the task of route planning and guidance, are not as well studied. Not only is the pilot faced with possible performance changes to the aircraft dynamics, but he or she is also tasked to create a plan of actions that will effectively take the aircraft down to a safe landing. In this plan, the many actions that the pilot can perform are closely intertwined with the trajectory of the aircraft, making it difficult to accurately predict the final outcome. Coupled with the vast number of potential actions to be taken, this problem may seem intractable. This is reflected in the lack of a pre-specified procedure capable of giving pilots the ability to find a resolution for this task.

This report summarizes a multi-year effort to examine methods to aid pilots in planning an approach and arrival to an airport following an aircraft systems failure. Ultimately, we hypothesize that automatic assistance to pilots can be provided in real-time in the form of improving pilot control of a damaged aircraft and providing pilots with procedural directives suitable for critical flight conditions; such systems may also benefit pilot training and procedure design. To achieve this result, a systematic, comprehensive research program was followed, building on prior research. This approach included a pencil-and-paper study with airline pilots examining methods of representing a flight route in an immediately understandable manner, and in a manner that would allow the pilot to modify an automatically-generated route and/or detect any inappropriate elements in an automatically-generated route. Likewise, a flight simulator study examined different cockpit systems for the relative merits of providing pilots with any of a variety of automated functions for emergency flight planning. The results provide specific guidance for the design of such systems.

Introduction

In training, pilots are taught the axiom Aviate, Navigate, and Communicate as a fallback procedure during times of crisis. Although useful in regards to providing a priority order for the tasks that need to be performed, such an sparse procedure does not cover in detail what the tasks are and when to perform them. As guidance during an emergency, students are taught several procedures governing such things as engine failures that specifically list out the tasks to perform. While these highly detailed procedures are helpful to the pilot in regaining stable control of the aircraft, they usually terminate with the simple instruction "Find a suitable landing area and land." At this point, the pilot is left to his or her own devices, with little guidance on how this goal should be achieved.

The subsequent (equally difficult and typically more lengthy) guidance task of planning and performing the maneuvers needed to land the aircraft has, in the past, relied heavily on pilot experience and judgement. No detailed written procedure covers this planning stage and creating a safe trajectory requires the pilot to know the effect that any action will have on the aircraft, given the environment and the status of the aircraft itself. For example, to turn onto the base leg of an approach, the pilot must realize the radius of the turn and initiate the bank ahead of time. In nominal flight, pilots already have the necessary "feel" for these relationships based on available information such as airspeed. However, in emergency situations where drastic maneuvers are performed, or where the aircraft performs differently, this inherent knowledge is no longer accurate and may even be detrimental if used improperly.

Unfamiliarity with the dynamics of an emergency situation places the aircraft at risk. Established training procedures for such things as medical emergencies, fires, and an assortment of control problems aim to reduce this unfamiliarity. However, many of these training procedures only specify certain aspects of the tasks in detail. For example, in case of a fire, pilots are asked to memorize the detailed procedures for the immediate aircraft systems management. The extent to which training and procedures are provided for the guidance task of establishing a safe trajectory down to the ground pales in comparison. Subsequently, training for this task is generally comprised simply of exposure to the situation in singular simulator scenarios during training. Although unstructured, this training is still useful in allowing pilots to form some heuristics about what plans may or may not work in similar emergencies. However, this lack of structure provides no guarantees, nor feedback, on the comprehensiveness and accuracy of the pilot's rules-of-thumb.

This lack of emphasis or structure on the guidance task during training may be a reflection of the vast amount of variation possible in the factors and circumstances that form an emergency. The inability to clearly fix all of the factors governing an emergency *a priori* makes it difficult to predict a suitable procedure or plan that would work in a wide range of cases. Thus, flight guidance is a difficult task, not only due to the complex calculations required to accurately predict a trajectory, but also due to the numerous possible plans available and the fact that in an emergency situation, the task needs to be carried out very quickly in stressful circumstances. Pilot formed heuristics serve to simplify this problem but impose assumptions on the aircraft system, such as envelope limits, and also assumptions about the surrounding environment, such as terrain, which may not be valid during an actual emergency. Thus, when such an event occurs in real life, it only shares a similarity with past training and therefore requires the pilot to extrapolate the required actions based on the perceived differences in the environment and aircraft dynamics. This, by its nature, is imprecise and a poor guarantee of safety.

This report summarizes a multi-year effort to examine methods to aid pilots in planning an approach and arrival to an airport following an aircraft systems failure. Ultimately, we hypothesize that automatic assistance to pilots can be provided in real-time in the form of improving pilot control of a damaged aircraft and providing pilots with procedural directives suitable for critical flight conditions; such systems may also benefit pilot training and procedure design. To achieve this result, a systematic, comprehensive research program was followed, building on prior research. This approach included a pencil-and-paper study with airline pilots examining methods of representing a flight route in an immediately understandable manner, and in a manner that would allow the pilot to modify an automatically-generated route and/or detect any inappropriate elements in an automatically-generated route. Likewise, a flight simulator study examined different cockpit systems for the relative merits of providing pilots with any of a variety of automated functions for emergency flight planning. The results provide specific guidance for the design of such systems.

Background and Previous Work

Cockpit Aids for In-Flight Re-Planning

Cockpit aids and safety systems have been widely proposed (and implemented). For example, the Traffic alert and Collision Avoidance System (TCAS) automatically detects and determines a plan of actions (or action?) to avoid nearby traffic. From time-critical planning systems such as TCAS, it is natural to extrapolate that there exists the possibility of automated planning for longer time-scopes, such as descent, arrival, approach and landing during emergencies, which could be expected to last anywhere from a few minutes to a half-hour.

However, the task of planning actions to guide the aircraft from its current position to a safe landing is a much more complicated problem due to the longer time-scales involved. The TCAS avoidance system can quickly determine the best action since it only involves the evaluation of an initial action followed by the resulting aircraft trajectory. In contrast, a pilot frames a longer-scale trajectory as a sequence of actions to be performed, such as turns, descents, and speed or power reductions. In this case, the initial actions and aircraft dynamics determine the trajectory only until another action is performed. To make things more complicated, as the aircraft trajectory is changed, so does the point at which subsequent actions are initiated; for example a turn to intercept a localizer will need to be initiated earlier or later in response to changes in the aircraft speed.

The complexity of the guidance task challenges the pilot to create a viable solution, especially in the context of a cockpit during a crisis. Numerous factors contribute to making this task difficult, from time-pressures to the unfamiliarity with new aircraft handling characteristics. Both pilot anecdotes and more theoretical studies of human machines systems identify that these stressors can force the pilot into more rudimentary modes of behavior in which pilots plan ahead with a lower time horizon, with less accuracy, based on less information, and with fewer goals.

Planning aids can be reasonably hypothesized to help with many aspects of the planning task. Rational planning can be broken down into (1) generation of alternatives, (2) imagining the consequences, (3) valuing (or evaluating) the consequences of the alternatives, and (4) choosing one alternative as a plan. Using this categorization, cockpit aids, to varying degrees, can be hypothesized to help with any or all of these elements of planning:

(1) Cockpit aids can potentially help pilots identify and generate alternatives for evaluation. For example, planning aids can highlight to the pilot items such as routes and destinations that do (or don't) meet obvious criteria such as fuel and time constraints, and may, with the addition of a knowledge database, be able to suggest useful first-cut solutions using basic heuristics tuned to the nature of the emergency (e.g., 'with pressurization problems, an immediate steep descent should be initiated.') However, for a cockpit system to have this capability requires it to have a significant level of intelligence about flight operations and about the goals and objectives of pilots during flight re-planning; this knowledge has not yet been identified and codified in a manner with sufficient resolution and suitable format for inclusion in a decision-aid.

(2) Given the computational complexity of predicting a trajectory as defined by discrete actions, the imaging consequences stage of planning can benefit from the assistance of a cockpit system capable of simulating the flight path resulting from a specified list of actions, which can then be displayed to the pilot and used in subsequent stages of planning. This element of re-planning has already been implemented in a previous study.

(3) Evaluating the consequences of any sequence of actions, likewise, can be assisted by a cockpit aid capable of highlighting safety risks and cost/performance measures of any particular trajectory. In engineering terms, this may be framed as giving the aid the ability to assess the viability and cost of a potential plan. However, implementation of this capability in a cockpit system means it again must be given substantial knowledge of the safety constraints and measures which a trajectory must satisfy, as well as detailed knowledge of the pilot's objectives and considerations in evaluating a plan; this knowledge has also not yet been identified and codified in a manner with sufficient resolution and suitable format for inclusion in a decision-aid.

(4) Finally, a cockpit aid can help in the selection of a plan when it is capable not only of assessing the performance of any individual plan, but also of conducting a search towards the set of acceptable (or, ideally, optimal) plans. However, the development of a cockpit aid to perform this function faces two major hurdles: First, as above, the system must be given substantial knowledge of the safety constraints and measures, and detailed knowledge of the pilot's objectives and

considerations in evaluating a plan; Second, the computational algorithms and method for searching through the intricate dynamics governing trajectories must be implemented. These hurdles are substantial: as with the other steps in planning, the knowledge has not yet been identified and codified in a manner with sufficient resolution and suitable format for inclusion in a decision-aid; and the development of computational algorithms and methods suitable for a trajectory governed by continuous-time aircraft dynamics interposed with discrete trajectory changes pose a challenging research problem.

Creating a system capable of assisting the pilot with these aspects of in-flight re-planning dictates several key technologies, as shown in Table 1. The primary technology is the ability to predict the causal relationship between the actions the pilot wishes to perform and the resulting aircraft trajectory. Such ability requires knowledge of the current and future flight characteristics of the aircraft, which may be degraded. In conjunction with resolving a model for the aircraft, a method is required to simulate the behavior of the combined pilot-aircraft behavior as the actions in a flight-plan are implemented. As prerequisites for system identification, other supporting technologies such as sensor design, fault diagnosis, and aircraft health systems are also needed. To provide for effective interaction with pilots, the planner will also require some study into the many possible different cockpit interfaces. Finally, to help with stage (4) of planning, the system would require some mechanism for generating a plan or, possibly, optimizing a flight plan.

Table 1 - Key Technologies Required

- [x] A method to predict the flight-path due to the causal relationship between actions and trajectory.
- [ ] System identification of the flight model.
- [x] Modeling of pilot-aircraft behavior.
- [x] Pilot-planner interface.
- [ ] Sensors, fault diagnosis, and aircraft health systems.
- [ ] Plan optimization (possibly)

Checked items denote focused areas of prototype.

Preliminary Work: Proof-of-Concept Development to Date

While it is understood that all of these enabling technologies are required to field a fully operational in-flight planner, a proof-of-concept first needs to be established before any specific technology is researched and matured. Therefore, research to date has implemented a proof-of-concept prototype called the Emergency Flight Planner (EFP). For the creation of this prototype, not all of the key technologies in Table 1 needed to be fully developed. Those that are unchecked can be emulated by their end-effect. At its core, the prototype is based on a fast-time simulation. Using this technique, the EFP can simulate in detail the aircraft trajectory through the flight-plan entered by the pilot.

The EFP's pilot interface is shown in Figure 1. It provides both a means for the pilots to enter their flight-plan into the EFP, as well as a means to provide feedback about the resulting trajectory. Through a virtual Control Display Unit (CDU), they can enter a flight-plan using actions that mimic those naturally used during the descent and approach. Although the interface merits of the CDU have been debated, most glass-cockpit pilots are familiar with it. Once entered, this flight-plan drives an internal EFP simulation of the future trajectory, taking into account the actual capabilities of the aircraft. The graphical representation of the aircraft trajectory provides feedback to the pilot about the feasibility of the plan. In the future, more mature planners may utilize the cursor to directly modify action triggers and properties directly on the plan and vertical profile views, through pop-up menus or other graphical interfaces. Such a fully graphical input interface may be more natural, but issues with the ambiguity in cursor modality and training requirements need to be resolved.

Once a list of actions is entered into the EFP, the system can simulate the aircraft as it flies this flight-plan. The resultant trajectory is displayed on the north-up plan-view. Utilizing the same symbology as a navigation display, this view allows the pilot to see the effect of system performance degradations on turns and other trajectory portions. In addition to the plan view, a vertical profile view is provided in order for pilots to better visualize the descent actions entered into the list. Since the EFP internally simulates the aircraft dynamics as it flies through the flight-plan, other states of the aircraft are available. A query view was incorporated into the EFP to allow pilots access to this extra information about any point along the future trajectory. These include time, speed, vertical speed, pitch, bank, fuel states, and flap and gear configurations. The format resembling a primary flight display (PFD) was selected for its familiarity to the pilot.

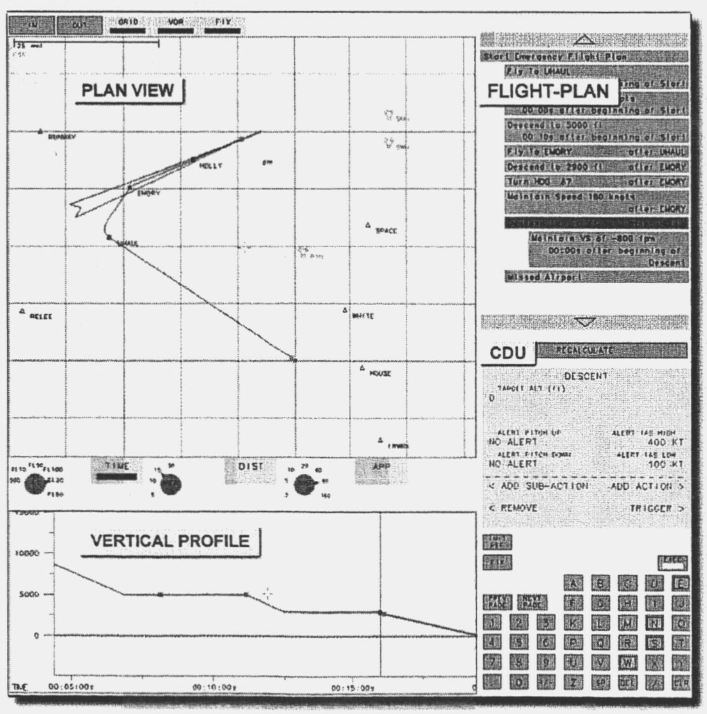


Figure 1 - EFP User Interface

(Black & White View of Color Display, Inverted for Clarity)

Preliminary Work: Experimental Evaluation of the Proof-of-Concept System

An experiment was conducted to provide an early evaluation of the efficacy of the EFP, with airline pilots as subjects. In addition to determining, quantitatively, whether the planner improves the ability of the pilot to perform a safe landing following a major system failure or emergency, the experiment also recorded the subjective opinions and comments of the pilots. In total, 72 runs were performed. Three different planning tools were made available to the pilots (charts only, basic EFP, and the EFP with a pre-loaded plan).

The results found that pilots were more favorable towards an automated planner that would present them with a pre-generated plan. This appeared to be mainly a factor of workload, since the basic EFP required substantial "programming" of the flight plan in order for it to generate a trajectory. Additionally, the performance data verified that having pre-generated plans from the EFP were better than plans created by the pilots using the planner because they could safely be accomplished in less time. This may be attributed to the general trend of the pilots in choosing the first plan that was viable, instead of iteratively refining the plan as originally expected; this tendency may be reasonably hypothesized to exist both with and without a cockpit aid, as pilots are likely to over-rely on the first plan they consider, even if it is not an optimal (or even good) plan.

Two insights into the need for an intelligent cockpit system were identified by the results. First, pilots' goals and objectives change with the context of different emergency situations. Second, pilots may not know enough about the aircraft to refrain from exceeding system limits – in fact, in some of the runs, the pilots were unable to capture their approach course without overly aggressive maneuvers. Together, these points illustrate the need for a cockpit aid – and that this aid must have a high degree of awareness of context and of pilot goals within that context.

Study 1: Presenting Pilots With Emergency Flight Plans

Although pilots often opted not to plan an emergency flight trajectory in the prior research studies, their responses on questionnaires indicated that they liked being provided with an emergency flight trajectory by an automated cockpit decision aid. In these studies, the trajectory was generated in real-time and represented as a procedure, whose elements included turns, speeds, vertical speeds, and aircraft configurations. This sentiment is also supported by similar studies (Funk, 1991; Layton, Smith, & McCoy, 1994; Smith, McCoy, & Layton, 1997), and leads to the question: How do we support a human in a task that is too hard for them to perform well in the time provided, but is too open-ended for automation to perform perfectly in every situation? We propose to address this question by looking at how we can help the pilots more effectively evaluate an emergency descent trajectory provided by automation and represented as an emergency procedure.

Previous research has shown that *procedure context* may aid in reducing the likelihood that a worker will blindly adhere to a computer-based procedure (Ockerman & Pritchett, 2000; Ockerman & Pritchett, 2004). In the previous research, many elements of procedure context were proposed from a literature review but only a few were empirically tested. Since the emergency descent trajectories can be represented to pilots as real-time procedures, the concept of procedure context may prove useful. The following sections provide a background of related research, a description of the study, and its results.

Background

Three categories of planning or replanning an aircraft's trajectory have been identified: strategic, tactical, and time-critical (Fan, Hyams, & Kuchar, 1998). Strategic planning is defined as being on the scale of hours and days. A strategic plan usually has one or two high level goals and specifies its actions abstractly, such as listing out airways to follow, without specifying in detail how the aircraft should be flown along the airways. Strategic planning is typically done for each flight to determine its flight path in the vertical, horizontal, and time dimensions. Strategic planning is typically done by the airline before flight in conjunction with air traffic control during the generation of normal operating flight plans.

Time-critical planning is typically on the order of seconds. A time-critical plan is made to meet one immediate, pressing constraint and has a high level of detail. These are usually evasive maneuvers to quickly avoid a collision with another aircraft or terrain. Currently, pilots are aided in time-critical planning with GPWS (Ground Proximity Warning System) and TCAS (Traffic Collision Avoidance System). Both of these systems warn of impending danger and suggest an immediate evasive maneuver to take in great detail such as specifying aircraft pitch, altitude and throttle settings.

The type of planning that this paper deals with, i.e., planning a trajectory down to landing in an emergency, is tactical planning. Tactical planning covers a future duration on the scale of minutes. A tactical plan has many goals and constraints and should be very detailed. Tactical planning is required when the situation makes the strategic plan unworkable; for example, pilots report that they most often have to do tactical replanning to avoid potentially dangerous weather systems (Fan et al., 1998). This type of planning or replanning is difficult because it requires the detail of time-critical planning but has a number of goals and constraints to consider. There has been almost no research into planning trajectories in emergencies and very little on tactical planning in general.

There are studies on more strategic flight planning that provide some insight into pilots' needs during tactical flight planning. Layton, Smith and McCoy (Layton et al., 1994; Smith et al., 1997) looked at en-route planning to avoid possibly troublesome weather systems. The experiments evaluated the interaction between an automatic weather planning aid and aviation users. The researchers determined that it would be infeasible to completely automate the task of planning a route around bad weather. The studies looked at three levels of aid being provided to the users. The lowest level of aid consisted of a 'sketching-only' system where the user planned the route from scratch and the system provided the results of various calculations (e.g., total distance, total time, and fuel consumption). The middle level of aid consisted of a 'route constraints and sketching' system where the user provides some constraints (e.g., shortest distance and lowest fuel consumption) and the system provides routes that meet the constraints. The highest level of aid consisted of an 'automatic route constraints, route constraints, and sketching' system where the system provided a route immediately. Overall, the users of the lowest level of aid developed more conservative plans than the users of the higher levels of aid accepted from the system. In addition, the users of the lowest level had difficulty in optimizing for a particular constraint, such as fuel efficiency. In the most difficult situations, users of the lowest level aid did not explore more alternatives than the users of the other two levels of aid though they did in the easier situations. Finally, users of the two higher levels of aid seemed biased toward selecting the computer-generated plan even when they generated their own plans as alternatives, thus showing a potential problem with automation-bias and/or over-reliance. This study illustrates the difficulty that arises when a task is difficult for the human to do on their own, but also too open-ended to always provide an optimal automated solution.

Another study that provides some insight into pilots' needs from an aiding system examined a cockpit task management aid (Funk, 1991; Chou, Madhavan, & Funk, 1996). In low-fidelity simulator studies they provided pilots with a color-coded task list of the tasks that needed to be accomplished for the current phase of flight. The task list was generated with consideration of the current phase of flight and state of the aircraft. The task management aid did improve the performance of the pilots in this study. In addition, the pilots reported that they liked to have an automatically generated list of tasks to be completed. However, this study did not examine what the pilots would do if the task list were incorrect in some way.

Finally, the previous work by the principal investigator examined the use of a prototype emergency flight planner (EFP) by airline pilots (Chen & Pritchett, 2001, detailed earlier). Of interest here, in one scenario the automatic EFP plan was inaccurate, but a majority of the pilots still followed it until it became obvious that it was based on a faulty model of the aircraft dynamics and would not lead to a safe landing. Likewise, pilots' interaction with automation and its representation of the trajectory was found to be problematic. (Pritchett et al., 2001)

All of these previous studies provide some support for the idea that pilots can be aided by the provision of an automatically-generated plan, represented as a detailed procedure; however, two studies (Layton et al., 1994; Chen & Pritchett, 2001) also indicate that pilots may have problems with rejecting a procedure that is faulty, a form of automation bias (Mosier & Skitka, 1999). Layton, Smith and McCoy (1994) suggest avoiding this problem by providing the pilot with more than one procedure and "forceing" them to choose between them or come up with their own. This may be a way to get the pilot more actively involved in the judgment process but, depending on the circumstances, it may be difficult to come up with two good procedures and, if one is bad, then the pilot's time during an emergency is diverted to an unnecessary task. In addition, current evidence suggests that the pilots would likely choose one of the procedures instead of creating their own. Another suggestion is to have a critiquing system (Guerlain, Smith, Obradovich, Rudmann, Strohm, Smith, Svirbely & Sachs, 1999) (based on an expert system) that watches the pilot's planning and lets the pilot know when they are doing something that the critiquer is not expecting. This suggestion has particular problems associated with it for tactical planning, which pilots normally do not explicitly perform and for which expert systems are not available.

Designers of decision aids often find themselves caught in an ambiguous situation. If they allow the operators to make their own decision, it may not be the best decision, particularly if time pressures are involved. Sometimes the dynamics of the situation cannot be reliably determined by a person without significant effort and training. However, if they allow the decision aid to make the decision, then it can be brittle in the face of unexpected circumstances. The choices are to either aid the operator in making a decision or present the operator with a finalized decision and aid the operator in evaluating that solution. The earlier studies in emergency flight planning suggest that the latter choice is more appropriate for this task.

However, previous research indicates that there will almost always be bias present if the system suggests a solution before the human has figured out their own solution, but, as mentioned earlier, it is not always feasible to wait or to expect a human to develop their own solution in the time provided. We hypothesize that providing an emergency descent trajectory that is supported by the presentation of *procedure context* may help the pilot to make a better judgment of its accuracy and feasibility.

Procedure Context

Borrowing the definition from the Merriam-Webster dictionary for context, procedure context is defined as procedure information that provides the interrelated conditions in which an individual procedure step exists or occurs. In other words, procedure context is procedure information that clarifies the meaning, purpose, conditions, and relationships of the individual procedure steps to the procedure, task, and environment. Procedure context situates the present procedure step in the larger purpose and organization of the procedure. Thus, the user is aware of more than just a series of isolated commands. Procedure context might also provide a window on the thought processes of the procedure designer and thus insights on how the procedure applies to the current situation. In this research, automation is the procedure designer.

Specific elements of procedure context have been identified through a review of literature and procedures (Ockerman & Pritchett, 2000). These elements may not comprise an exhaustive list but do provide a sufficient basis to begin to evaluate the possible benefits of procedure context in different environments (Ockerman & Pritchett, 2004). These individual elements are either known early in the procedure's life-cycle or are implicitly part of a completed procedure; as such, procedure context is not adding information to a procedure but making the relevant pieces of information explicit to the user.

It is typically not possible to provide procedures that meet the needs of every possible situation (Wright, Pocock, & Fields, 1998; Vicente, 1999). As such, procedures serve as guidelines to the way the task should be completed if the conditions match those for which the procedure is intended. Thus it is important to encourage designers to consider users as 'cognitive beings' as opposed to servants of an aiding system. As a cognitive being, the user can react to the situation and interpret the procedure to fit the current situation (Suchman, 1987; Dien, Montmayeul, & Beltranda, 1991; Wright et al., 1998).

Some of the over-reliance literature points to the importance of communication of actions and motives in the appropriate interaction of humans and machines. Although one cannot force a user to be cognitively involved in a task, without sufficient information on the underlying purpose and structure of an aid informed cognitive involvement is not possible. Without sufficient information to make decisions, and to evaluate and assess the situation, the user is reduced to following the instructions one by one (Swezey, 1987). Therefore, procedure context is the information necessary to encourage users to be cognitive beings and thus appropriately rely on an aiding system.

The procedure context elements used in this study are considered explanatory, which means they supply meaning, purpose, relationships, or conditions. Explanatory elements are a window into the procedure designers' reasons for the current procedure, providing information about reasoning, conditions for use, and a strategy for using the procedure. The explanatory procedure context lets the user know when the procedure does not match the current situation and can aid the user in determining other ways of accomplishing the task in a different but safe manner.

Explanatory procedure context elements are frequently not included in a finished procedure but might be covered when a user is trained on the procedure or on the job. However, the designer of the procedure presumably knows explanatory procedure context elements since this type of information is needed in designing a procedure. The designer must convert the dynamics and structure of a task and its relationship to an external, changing environment into an understandable and effective static representation that can be used over and over. Unfortunately, all this knowledge is rarely explicitly published in the final procedure and is often effectively "lost" to the user. It has been noted that sometimes a significant amount of time transpires before a worker might realize why the procedure is designed as it is (Hutchins, 1995), and even then there is no guarantee that he or she has formed a correct interpretation.

Explanatory information about the task, such as rationale, has been shown to increase the rate at which each step is read, aids in recalling the steps of the procedure at a later time, and improves performance when performing the task without the procedure (Smith & Goodman, 1984). In addition, knowing the background of a procedure can ease worker resistance to a procedure by providing insight into the relevance of the procedure to effective task completion (McCarthy, Wright, Monk, & Watts, 1998; Wright et al., 1998; deBrito, 1999).

The three explanatory elements used in this study are:

<u>Rationale</u> The rationale element provides the reason for each procedural step; that is, why this step is included in the procedure. In one study, inspection of a commercial pilot's own copy of procedures found that 38% of the annotations made by the pilot could be classified as notes on "why a procedure is the way it is" (Wright et al., 1998, p. 3). Knowledge of a step's rationale can aid the user in understanding when this particular step is appropriate for the current situation and improve his or her knowledge about the procedure.

Triggering Conditions The triggering conditions element identifies the events that signal the starting or stopping of a procedural action. Triggering conditions are events that are external to the procedure. For example, crossing a fix may trigger a change in speed and descent rate. Making procedure followers aware of triggering conditions can help them to determine not only when a step should be done but also provide insight into the relationship between the environment and the task.

Ordinality The ordinality element defines any order requirements; i.e., when an action must be done before or after other actions, or when it does not matter. For example, an aircraft must be at a certain speed before deploying the flaps: otherwise the aircraft structure could be damaged. However, other actions may not have ordinality requirements and can be done in any order or simultaneously. Thus, the ordinality differs from the triggering conditions in that it only concerns the order of the procedure steps, not their relation to external events.

Method

This study investigated how a real-time procedure should be represented so that a pilot can quickly determine whether a procedure generated by an emergency flight planning aid will lead to a safe landing and therefore whether or not he or she should follow it.

Participants and Apparatus

A total of 32 active, commercial pilots participated in this study, with 28 of them volunteering demographic data. Those pilots had an average of 11,500 flight hours, with just over 4000 hours in glass cockpits. Twenty-two of the participants were captains, five were first officers, and one was neither.

The eight emergency scenarios they evaluated were presented on paper and consisted of 6 items (see Figure 1): (1) a description of the emergency that has occurred along with a picture of the primary flight display and navigation display indicating current aircraft state, (2) an enroute map for the new airport, (3) an approach plate for the new airport, (4) a STAR chart for the new airport, (5) horizontal and vertical map displays of the suggested descent path, and (6) a text procedure for the suggested descent.

Bruin Briefing

The spoilers were unexpectedly deployed. Now they are stuck. You have control of the aircraft and it appears to be stable. While testing the aircraft's maneuverability you have determined that you can bank the plane between the red bank limits on your PFD.

The closest available airport of sufficient length is Bruin International, on the Islands of Virginia. You start a descent towards the airport.

At this moment your navigation display is showing as below.

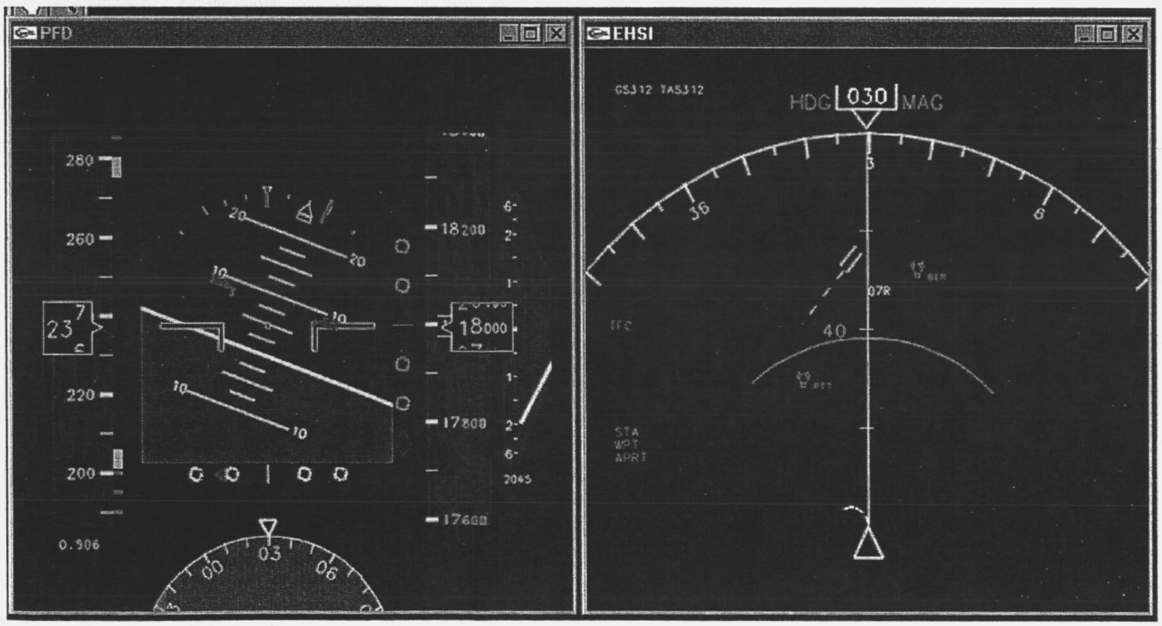
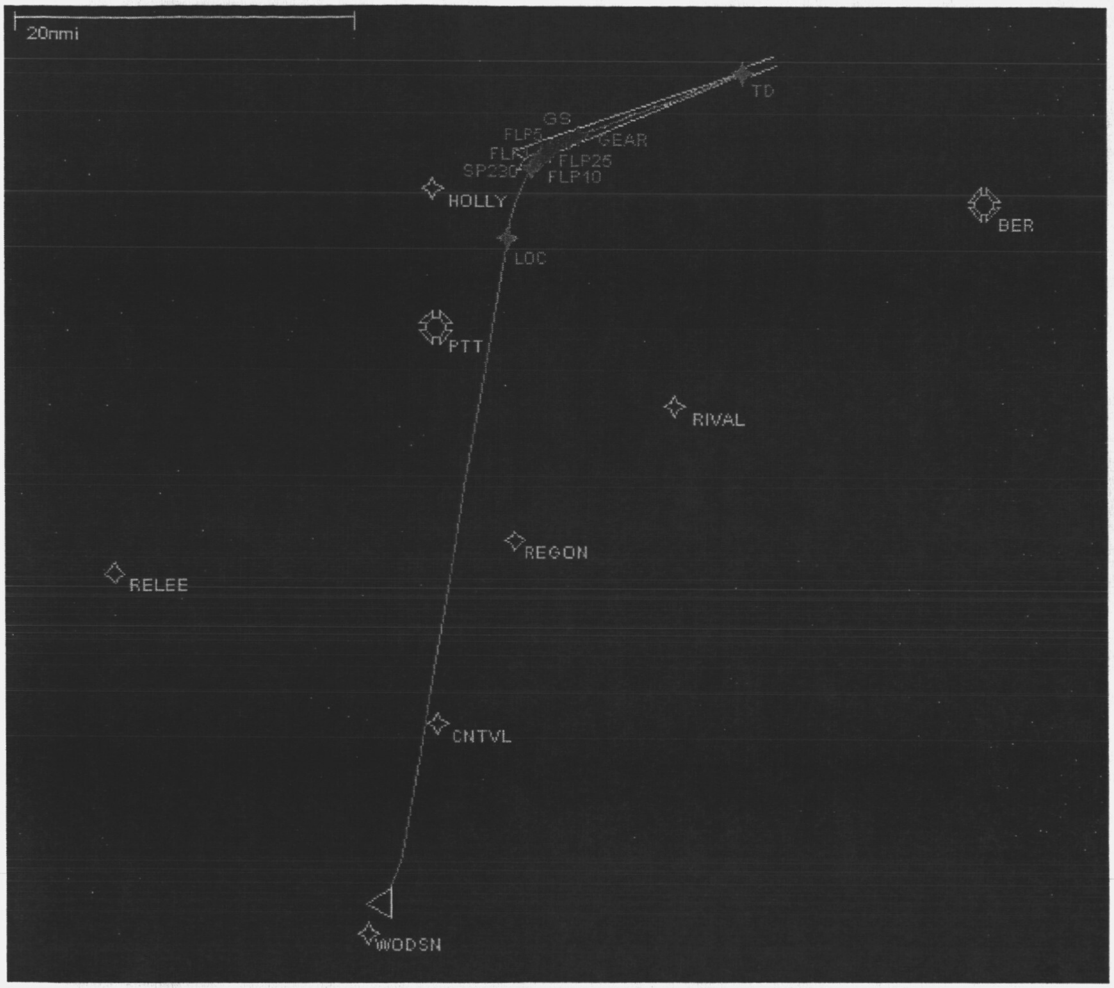


Figure 1a. ***Description of emergency with primary flight display and navigation display***

Bruin



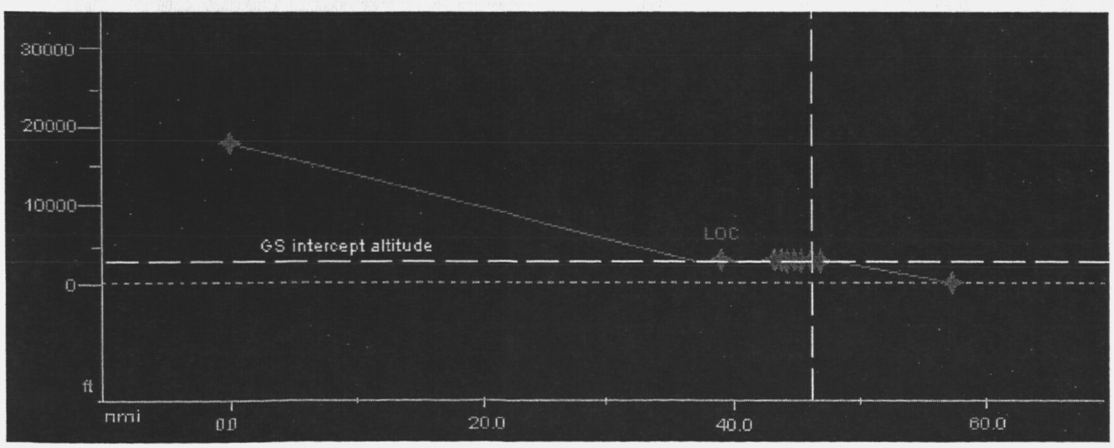


Figure 1b. ***Horizontal and vertical map displays***

Procedure

The pilots were told that they were the captains of a glass cockpit commercial jet and that an emergency had occurred which required that they land immediately at the nearest airport. The emergencies may or may not have affected the performance of the plane and none had terrain or traffic conflicts. Each scenario consisted of a written description of the emergency and the current location of the plane. To replicate the time-criticality of emergency situations, the pilots were given 3 minutes to evaluate each procedure, which contained actions that would establish a trajectory to the airport, and record their response on a questionnaire. On the questionnaire, the pilots categorized each procedure as one they would be comfortable flying or one they would not be comfortable flying, and explained why or why not. They also provided their confidence in their response as a percentage.

Design

This experiment used a 2^4 factorial design. The first factor was the condition of the aircraft (performance altered [PA] or not [NPA]), the second factor was the accuracy of the procedure, the third factor was the structure used in the presentation of the procedure, and the fourth factor was the presence of rationale (i.e., explanations).

Performance of the aircraft in each scenario was either altered in some way [PA] (e.g., lost engine or loose aileron control surface) or was not [NPA] (e.g., sick passenger). Determining the future flight trajectory of a performance-altered aircraft is more difficult due to its unpredictable nature and inexperience with an aircraft in that particular condition.

Half of the eight procedures were inaccurate. The inaccuracies were of two types. In one type of inaccuracy the graphic map display accompanying the procedure was redrawn to show a much tighter turn radius at some point in the descent than feasible for the aircraft's speed and configuration at that time. In the other type the graphic vertical profile was altered to show an infeasible glide slope intercept, i.e., where the aircraft was at least 1000 feet too high to intercept the glide slope with the descent rate given by the procedure. In both cases the text accurately listed a series of actions that created the infeasible procedure.

The two structure variants and two presence of rationale variants resulted in four distinct display formats The structure was either sequential or concurrent. The sequential structure listed all the actions that were required to complete the descent in a single column and noted when to do each action by attaching a 'fix' to the action that was also presented on the graphical display (see Figure 2). The concurrent structure listed the actions in a matrix where the columns related to horizontal motion, vertical motion, speed, or configuration, with all concurrent actions listed in the same row (see Figure 3). Again each row was notated with a fix and/or event to indicate when they should be done. The rationales, when provided, explained why an action should be done in general and/or done at a particular time (see Figure 2). Thus, the four formats are sequential, sequential with rationales, concurrent, and concurrent with rationales.

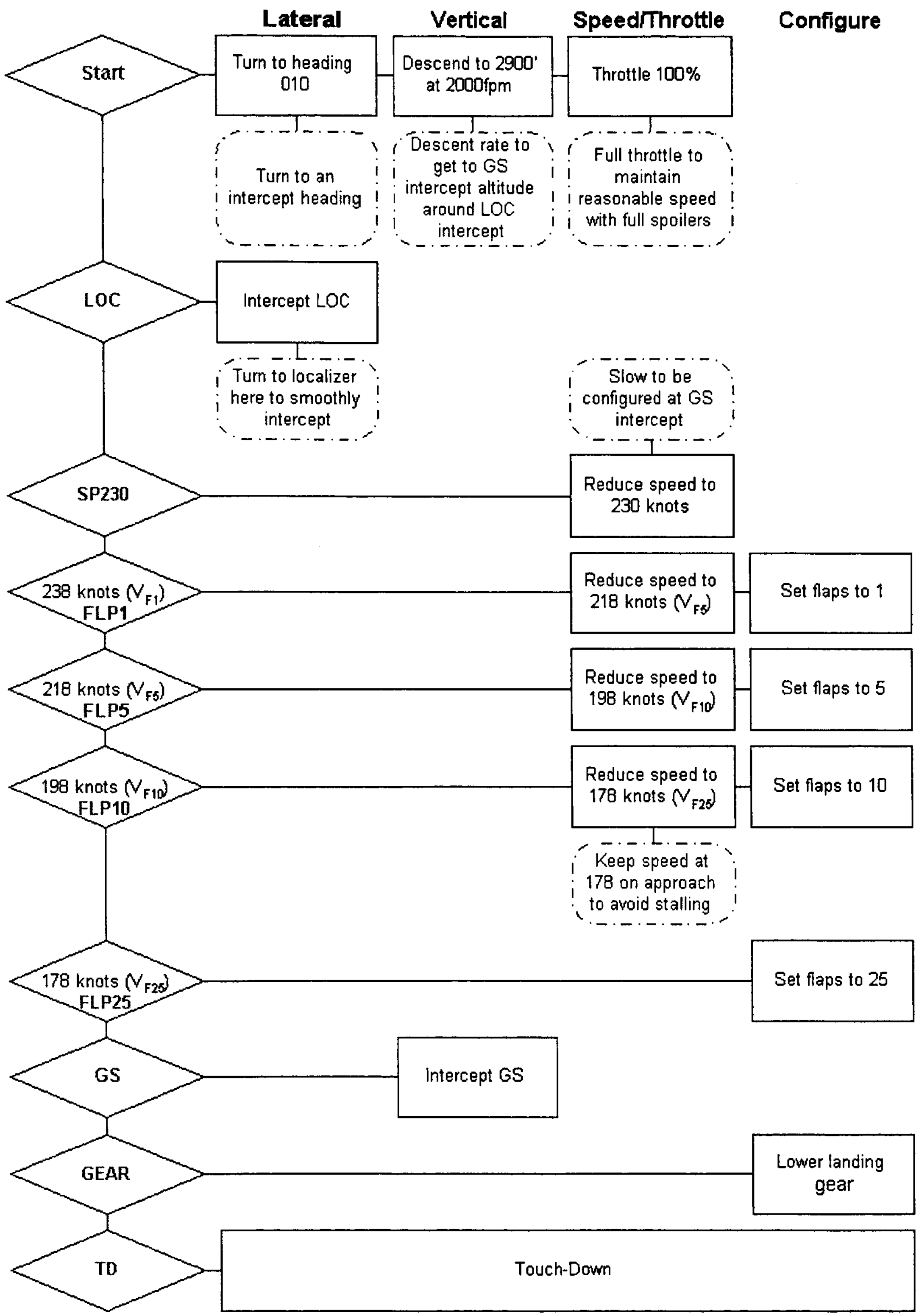


Figure 2. ***Sequential structure with rationale***

Homepark

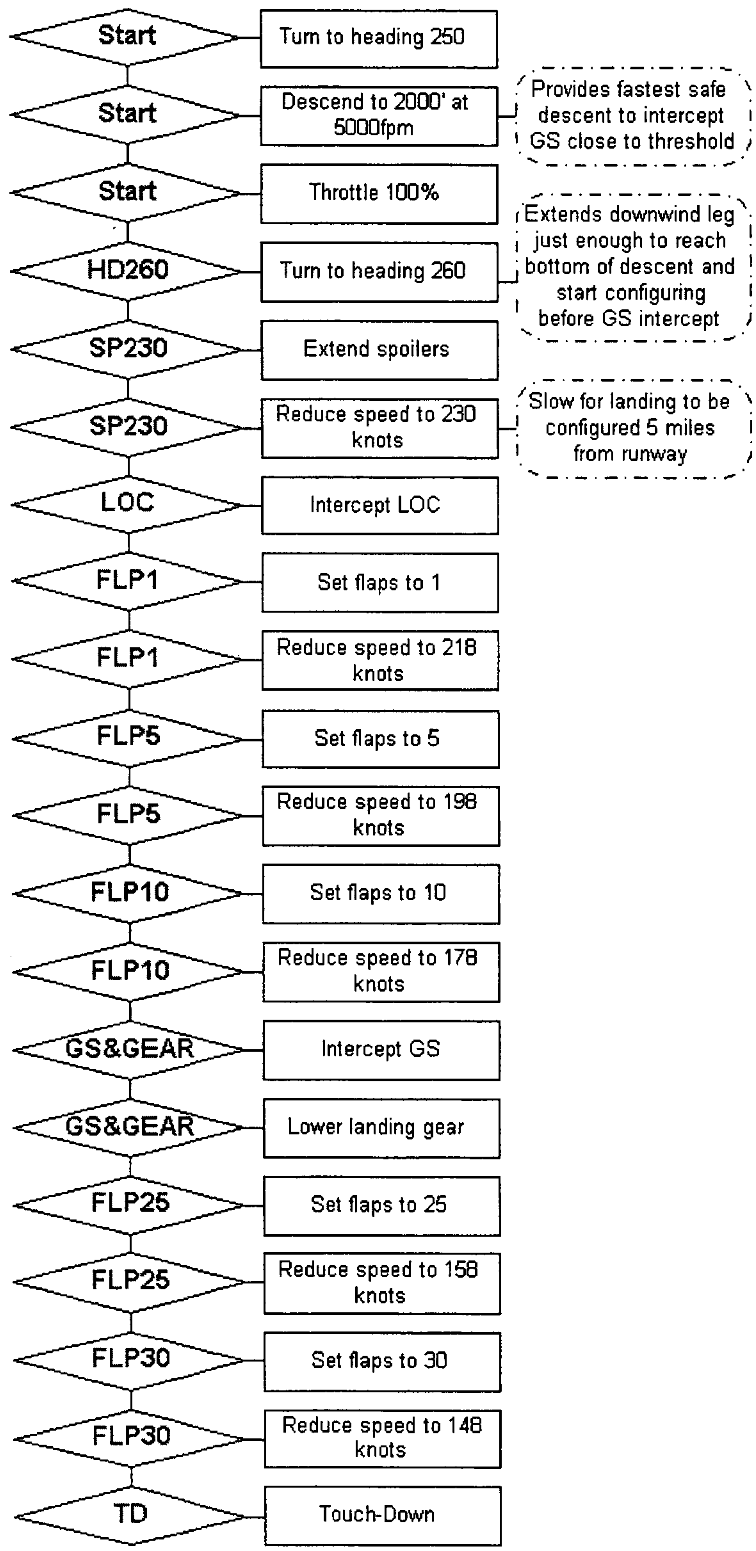


Figure 3. ***Concurrent structure without rationale***

Table 1. **Scenario Descriptions**

Scenario	Condition	Accuracy	Structure	*Rationale*
1	PA	Accurate	Sequential	Not present
2	PA	Inaccurate	Concurrent	Not present
3	NPA	Inaccurate	Sequential	Not present
4	NPA	Accurate	Concurrent	Not present
5	NPA	Inaccurate	Concurrent	Present
6	NPA	Accurate	Sequential	Present
7	PA	Accurate	Concurrent	Present
8	PA	Inaccurate	Sequential	Present

We blocked on the factor rationale to mitigate learning effects, so the pilots either saw a combination of scenarios 1-4 and then scenarios 5-8 or they saw scenarios 5-8 and then scenarios 1-4. We used 8 different scenario orders; four pilots did each order of scenarios (see Table 1).

These display formats correlate with the three procedure context elements discussed earlier. The rationales used as an experimental factor map directly to the rationale procedure context element. The trigger condition and ordinality elements map to the plan structure. The trigger conditions for the sequential structure are listed as fixes while the trigger conditions for the concurrent structure are fixes and speed events. For example, at the point where the aircraft will have slowed to a safe extension of flaps 1, the trigger condition is fix FLP1 and the speed event 238knt (V_{F1}). The ordinality element in the sequential structure is a simple linear order that arbitrarily orders actions at the same fix. For the concurrent structure the ordinality element is more clearly shown, with the matrix format illustrating that actions attached to the same fix should be done at roughly the same time but that the exact order is not important.

Measurements

Measurement consisted of the pilots' responses to the presented procedures and a follow-up questionnaire. These included the pilots' responses (i.e., whether they would or would not follow the procedure), the confidence they assigned to their response, the correctness of their responses (i.e., whether their response matched the flight procedures' accuracy), and the correctness of the pilots' reasoning about the procedure as recorded in written comments. The questionnaire measurements are the pilots' opinions about the different presentations of the procedure.

Results

There were 256 data points for each of the response variables: pilot responses, pilot confidence, correctness of pilot responses, and correctness of pilot reasoning. In addition to the four experimental factors, the pilot group, which represents the order in which the pilots saw the different scenarios, was also examined for main effects, but was shown to not have an effect for any of the response variables.

Pilot Responses

An analysis of variance (ANOVA) general linear model (GLM) (type III adjusted sum of squares) was used to analyze for effects. The GLM for pilot responses versus the four experimental factors: performance of the aircraft (PA or NPA), procedure accuracy (accurate or inaccurate), procedure structure (sequential or concurrent), and the presence of rationale showed that only the aircraft condition, PA or NPA, is a statistically significant factor (see Table 2). Examination of the data shows that pilots were more likely to respond 'No' (not comfortable following) in performance-altered conditions and 'Yes' (comfortable) in non-performance-altered conditions.

Table 2. **GLM for Pilot Response Versus the Four Experimental Factors**

Source	*DF*	*Seq SS*	*Adj SS*	*Adj MS*	*F*	*P*
Condition	*1*	*4.0000*	*4.0000*	*4.0000*	*17.00*	*0.000*
Accuracy	*1*	*0.0625*	*0.0625*	*0.0625*	*0.27*	*0.607*
Structure	*1*	*0.5625*	*0.5625*	*0.5625*	*2.39*	*0.123*
Rationale	*1*	*0.2500*	*0.2500*	*0.2500*	*1.06*	*0.304*
Error	*251*	*59.0625*	*59.0625*	*0.2353*		
Total	*255*	*63.9375*				

Pilot Confidence Level in Response

The GLM for pilot confidence level versus the experimental design factors also had a significant factor – performance of the aircraft once again (see Table 3). Examination of the data showed that the pilots had a higher level of confidence with non-performance altering conditions. This is not surprising but does show that the pilots did account for the aircraft performance when making their judgment.

Table 3. **GLM for Pilot Confidence Versus the Four Experimental Factors**

Source	*DF*	*Seq SS*	*Adj SS*	*Adj MS*	*F*	*P*
Condition	*1*	*1891.6*	*1896.1*	*1896.1*	*4.15*	*0.043*
Accuracy	*1*	*100.0*	*97.1*	*97.1*	*0.21*	*0.645*
Structure	*1*	*19.0*	*19.4*	*19.4*	*0.04*	*0.837*
Rationale	*1*	*113.4*	*113.4*	*113.4*	*0.25*	*0.619*
Error	*248*	*113425.7*	*113425.7*	*457.4*		
Total	*252*	*115549.6*				

Correctness of Pilot Response

For the correctness of the pilots' responses when compared to the accuracy of the presented procedures, none of the four experimental factors had a statistically significant effect. In fact, on the whole the pilots did little better than chance (52%) on correctly judging the accuracy of the presented procedures.

Correctness of Pilot Reasoning

Finally, the correctness of the pilots' reasoning for accepting or not accepting a procedure was analyzed by categorizing pilots' reasoning and then comparing these categorizations with those provided by a subject matter expert. Looking at the four experimental factors versus correctness in pilot reasoning showed that two of the factors were statistically significant: accuracy of the presented procedure and rationale (see Table 4). Examination of the data showed that the pilots had more accurate reasoning for accurate scenarios. This is not surprising since they basically only had to agree that it was done correctly. In addition, further analysis showed that procedures displaying rationale resulted in more correct reasoning by the pilots.

Table 4. **GLM for Pilot Rationale Accuracy Versus the Four Experimental Factors**

Source	*DF*	*Seq SS*	*Adj SS*	*Adj MS*	*F*	*P*
Condition	*1*	*0.0352*	*0.0352*	*0.0352*	*0.21*	*0.650*
Accuracy	*1*	*0.6602*	*0.6602*	*0.6602*	*3.87*	*0.050*
Structure	*1*	*0.1914*	*0.1914*	*0.1914*	*1.12*	*0.290*
Rationale	*1*	*1.7227*	*1.7227*	*1.7227*	*10.10*	*0.002*
Error	*251*	*42.7930*	*42.7930*	*0.1705*		
Total	*255*	*45.4023*				

Questionnaire Results

The questionnaire measures came from the opinions of the pilots on the four different formats. Of the four different formats, 45% of the pilots with an opinion preferred the concurrent with rationale format. Overall, 67% of the pilots with an expressed opinion preferred the concurrent format over the sequential format, and 91% of the pilots with an expressed opinion preferred having rationales over not having rationales.

Discussion

As is often the case with experiments of this type, There were large individual differences between the pilots in their acceptance of the procedures and their reasoning for that acceptance. In addition, overall the pilots were little better than chance at distinguishing procedures that correspond to a trajectory leading to a safe landing from those that do not. This is not surprising since this is a task that they have rarely, if ever, performed and does not have any standardized training. Pilots do practice emergency situations in simulator training but these focus more on the initial procedural response to the emergency than on planning a new trajectory on the fly to descend to an airport.

However, there are several interesting aspects of the results of this study. Not only were the pilots more likely to accept a procedure in the NPA condition, they also had more confidence in that acceptance. This may be due to a higher level of comfort with a "normal" aircraft that should perform as expected. However, this comfort may be misapplied, as they often indicated they would follow inaccurate NPA procedures.

When the four experimental factors were examined in relation to correctness of pilot reasoning, procedure accuracy and the presence of rationale were significant. Having the correct reasoning for an accurate procedure was not overly difficult since basically the pilot had to just accept the procedure as correct without listing caveats. More interestingly, the procedures with rationales lead to a more correct reasoning by the pilot for acceptance or non-acceptance of a procedure. The pilots also reported that they liked being provided with the rationale of a procedure. There is no support for the structure of the procedure impacting the pilots' responses or correctness in the objective results, although a majority of the pilots did prefer the concurrent structure over the sequential structure.

This study suggests that the presence of rationales or explanations for automatically generated decisions can aid the operators in more correct reasoning about that decision; however, it did not impact the correctness of their response to follow or not follow the decision. Further investigation is needed to understand both why pilot performance at this task is so poor and why this contextual information did not improve the pilots' judgments. However, these results indicate that including rationale with a suggested plan of action can improve some aspects of operator performance which might lead to enhanced system function and help operators deal with the potential brittleness of automatic systems. This finding has implications for both the design of procedures and the design of automatic systems that suggest courses of action to a human in situations that cannot be fully evaluated by the procedure or automatic system.

Study 2: Automated Systems and Emergency Flight Planning

Research on planning has emphasized automation with a view to alleviating the workload on pilots and dispatchers either by automating planning processes or delegating decision-making away from the flight deck. However, few studies have examined the behavioral aspects of planning in general and the impact of automation in particular. This is especially true for "tactical planning", i.e., planning in a time horizon on the order of tens of minutes. Thus it is hypothesized that some automation in the flight deck should be available that could assist air transport pilots in tactical planning by considering many factors about the immediate and near-term situation.

In this study, a flight simulator experiment was conducted to study airline pilot performance in tactical replanning tasks using several different autoflight systems. Each pilot was placed into either a non nominal or an emergency situation which required replanning. All pertinent checklists were assumed to have been performed and the aircraft was currently in stable flight. His or her immediate task was to replan the current flight and fly down to the final approach.

Background And Motivation

Formally, a flight plan is a list of destinations or waypoints, their associated altitudes and speeds, and a destination which is to be filed with a legal authority before a flight. Functionally, the term "plan" can also refer to a succession of goals and actions that are designed and executed to fulfill the final objective.

In air transport, flight plans are typically created by the pilot and dispatcher, and approved (and potentially modified) by air traffic operators before take-off. In addition, a substantial amount of re-planning may need to be done on the fly during flight, where pilots have real time access to more current information sources.

Flight planning is essential as it is a process by which a suitable set of high level actions is created that will enable the flight to reach its destination. At a base level, flying an aircraft is essentially an exercise in managing available resources including time, fuel, energy, or a combination thereof. Management of these resources is crucial to an efficient flight and to do this the pilot must incorporate knowledge about the current environment. In higher workload situations, especially emergencies, pilots may face near impossible demands on their time. Flight planning offers a reduction of workload during later stages by enabling the pilot to follow a predetermined plan, and also can establish an efficient and safe trajectory throughout the flight.

Technological developments have made it possible to automate more and more functions in the flight deck and in other high workload and dynamic domains. Automation in the flight deck has evolved from the most basic autopilots to sophisticated systems such as flight management systems. Similarly, automation to maintain flight safety has also seen a sea change with the development of systems such as the Traffic Alert and Collision Avoidance System (TCAS) and the Ground Proximity Warning System (GPWS). The introduction of advanced technology on modern flight decks has succeeded in increasing the precision and efficiency of flight operations.

As part of this trend, systems have been developed that assist pilots with time-critical planning. For example, TCAS calculates an avoidance maneuver and displays it to the pilot, and the GPWS has a built-in aural alert which alerts the pilot to perform a standard avoidance maneuver. Due to their time critical nature, such re-planning tools have the characteristic of a forcing function on the pilot and are inherently automatic and assertive in nature. Another important element of modern flight deck automation is the Flight Management System (FMS).

"The FMS supports the pilots in a variety of tasks, such as flight planning, navigation and guidance, performance management and monitoring of flight progress." (Sarter and Woods, 1994). The major FMS interfaces for the pilot are the mode control panel (MCP) and the control display units (CDUs).The FMS is also intricately tied to many cockpit displays, including the primary flight displays (PFDs), and electronic horizontal situation indicators (EHSI), which display information about the autoflight modes and the current route of flight.

The CDUs consist of a keyboard and a data display screen. The keyboard is used by the pilots to enter data that define a flight path and to access flight related data available in the numerous display pages. The pilot-entered flight path is continuously updated to reflect current flight status and is presented on the EHSI when in map mode. This allows pilots to monitor progress along the path. In the EHSI plan mode, the pilot can visually check modifications to the active flight plan.

The MCP is used to activate different automatic flight modes such as: Vertical Navigation (VNAV), Lateral Navigation (LNAV), Heading Select (HDG SEL) and Flight Level Change (FLCH). The pilot can also use knobs on the MCP to dial in targets for individual parameters (airspeed, heading, altitude, and vertical speed), which are tracked when their corresponding automatic flight mode is activated. To find out which FMS modes are currently active, the pilot can monitor the flight mode annunciations on the PFD. These provide data on the active (or armed) pitch and roll modes and on the status of the autopilot(s). They also indicate the status and mode of the autothrottles, which can be set to either manual or automatic mode for speed and altitude control. The various FMS interfaces combine to provide the pilot with a high degree of flexibility in selecting and combining levels of automation to respond to different situations.

The FMS can also help with flight planning. When the authorized flight plan is being entered into the FMS while the aircraft is at the gate, it would be considered as being used for strategic planning purposes, and when a reroute is being planned in the air for the next few minutes of flight, it would be considered as being used for tactical planning. The FMS can also provide a "what-if" capability (Honeywell, 1996). For example, the pilot can query the FMS to determine how much extra fuel will be burned if he or she increases speed by Mach 0.02. This provides pilots the information needed to evaluate new plans.

Recent accidents and incidents involving glass aircraft suggests that the increase in automation in the flight deck also have a degree of operational burden associated with them. This can lead to various breakdowns in the overall human-machine system. This has been hypothesized to arise from the complexity of the FMS itself and/or poor portrayal to the pilot of its functioning. Studies exploring the pilots' mode awareness and understanding of the functional structure of automation are plentiful. However, less research has examined its utility for tactical planning.

Objectives

The main objectives of this experiment were to study:

- Pilot planning performance at in-flight re-planning in non-nominal and emergency flight conditions;
- Pilot planning behavior for in-flight re-planning in non-nominal and emergency flight conditions;
- The impact of cockpit automation on the planning process.

Additionally, this experiment was also a preliminary investigation of an intelligent cockpit aid capable of automatic flight plan generation. This investigation was preliminary in that only the concept of such a system was explored and the plans used for the experiment were preprogrammed into the planning interfaces.

Method

In each experiment the pilots faced either a non nominal or an emergency situation about 30 minutes (85-90 miles) from landing. Before that start of each flight, pilots were given a scenario briefing (Appendix B.4) along with paper charts. They were given 25 seconds to go through the charts before the run was started. Their task was to replan the route while in flight, with the assumption that the all pertinent checklists had already been completed, the situation contained, and control of the aircraft had been regained just before the run started.

A confederate pilot was present in all runs. The main function of the confederate pilot was to get clearances from air traffic control, deploy the flaps and gear when asked by the test pilot, and to enforce the type of automation used for the run, i.e., in the CDU (and its variants) cases, pilots were not allowed to use the MCP and vice versa.

Sixteen pilots took part in the experiment. Each pilot ran nine flights for a total of 144 runs. The run order was determined by a test matrix which was a balanced combination of two independent variables: type of automation and scenario type, based on a Latin Square design. The types of automation tested were MCP, CDU, CDU+ and CDU++. The scenario types were classified into two types, non nominal and emergency.

The simulator logged important data including aircraft state variables (such as speed, distance, latitude and longitude) and identifiable actions in the autoflight systems (such as speed changes, altitude changes and heading changes). Additionally, pilots were also asked to fill a questionnaire at the end of each run and at the end of the experiment.

Scenario Design

To avoid pilot familiarity with a common arrival route, fictitious airports and arrival routes were used for the experiment. The airports were adapted from those previously utilized in two other experiments to study arrival procedures and cockpit display of traffic information (Yankosky and Pritchett, 1999) and the Emergency Flight Planner (EFP) (Chen and Pritchett, 2000). A new airport for the training runs and a number of waypoints, fixes and navigation aids were added to the existing charts. Terrain was not a consideration in the experiment.

A total of ten airports and their related charts were used for the experiment, one for each scenario. Four airports were reserved for non nominal scenarios, four for emergency scenarios, one for the faulty Autoplan scenario and one for the training scenario. The tenth airport reserved for the faulty Autoplan scenario and was used for both non nominal and emergency scenarios.

All the scenarios were designed to be of equal difficulty. The initial positions of the aircraft at the start of the scenarios were placed such that pilots could choose to approach the airport from either the left or the right of the runway. The run was terminated once pilots had intercepted the localizer at glideslope altitude at the outer marker. Before the start of each run, pilots were given a briefing sheet. Given below is a sample of a non-nominal scenario briefing and an emergency scenario briefing.

Sample non-nominal briefing:

Atlantic Briefing

You are heading along the Townhouse One Arrival at Atlantic International Airport and are 13 miles past VOR CLR[114.0 CLR], when you receive word from ATC that there is severe turbulence directly in your path ahead and spanning the area shown in your en-route chart.

The destination is runway RW29L at Atlantic International. Your current state is:

- heading 347°
- 13000 ft altitude (-1200 fpm)
- 290 IAS

Start your replanning from this point.

Sample emergency briefing:

Bruin Briefing

You were heading along the Braddock Arrival, when your alarm systems detected a fire in the cargo hold. The fire has been put out by the flight attendants, but the extent of the damage is not clear. You are 52 miles past VOR BRN [114.0 BRN], by the time you decide to declare an emergency and all standard procedures and checklists have been completed.

The destination is runway RW18R at Bruin International Airport and your current state is:

- heading 34°
- 9000 ft altitude (-1200 fpm)
- 250 IAS

Start your replanning from this point.

Additionally, pilots were also provided paper charts for the area based on the current Jeppesen standard. The paper charts included an en-route chart, a Standard Terminal Arrival Route (STAR) chart and an approach plate. These charts are shown in Figures 4, 5 and 6.

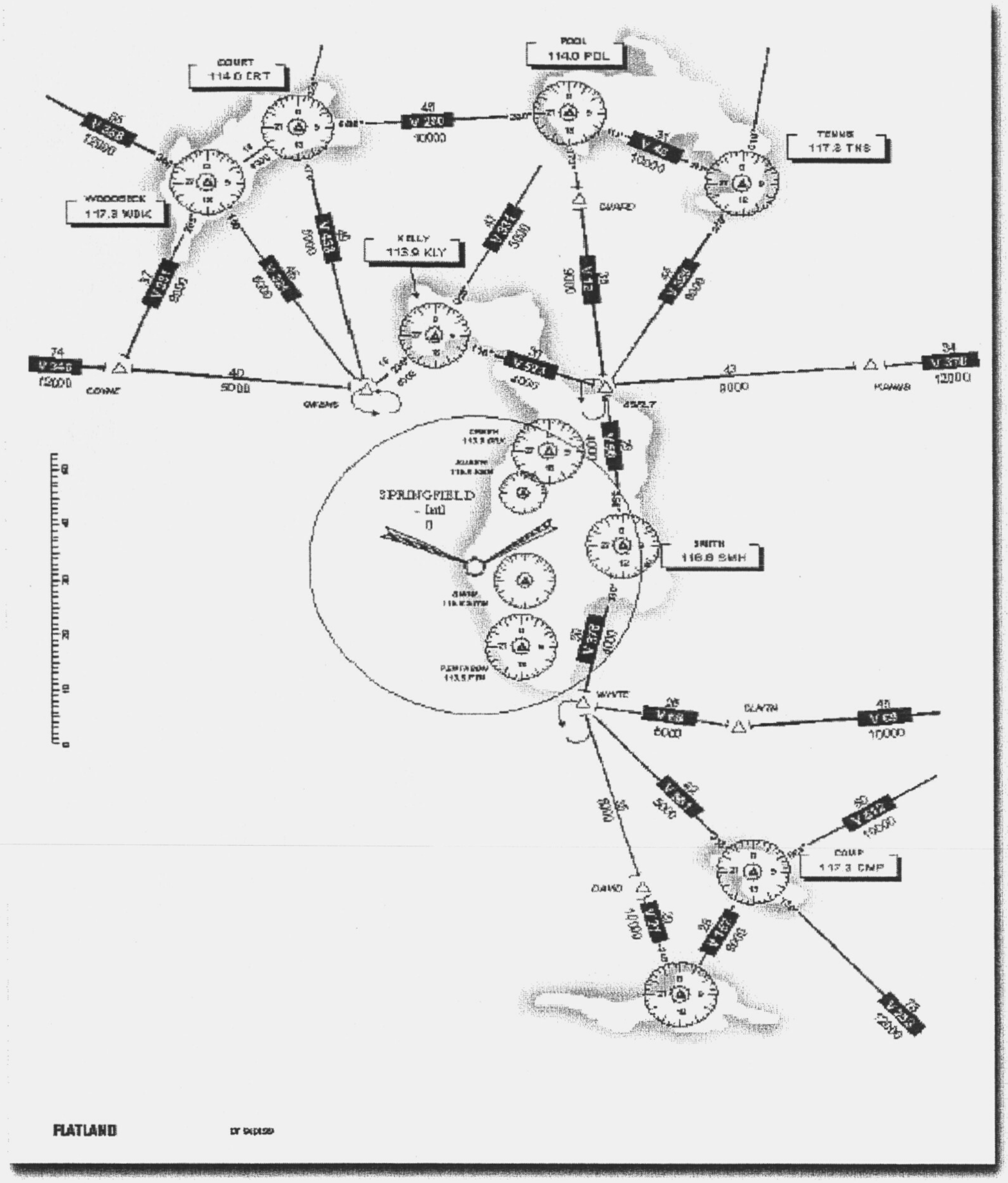


Figure 4 - Sample En-route Chart

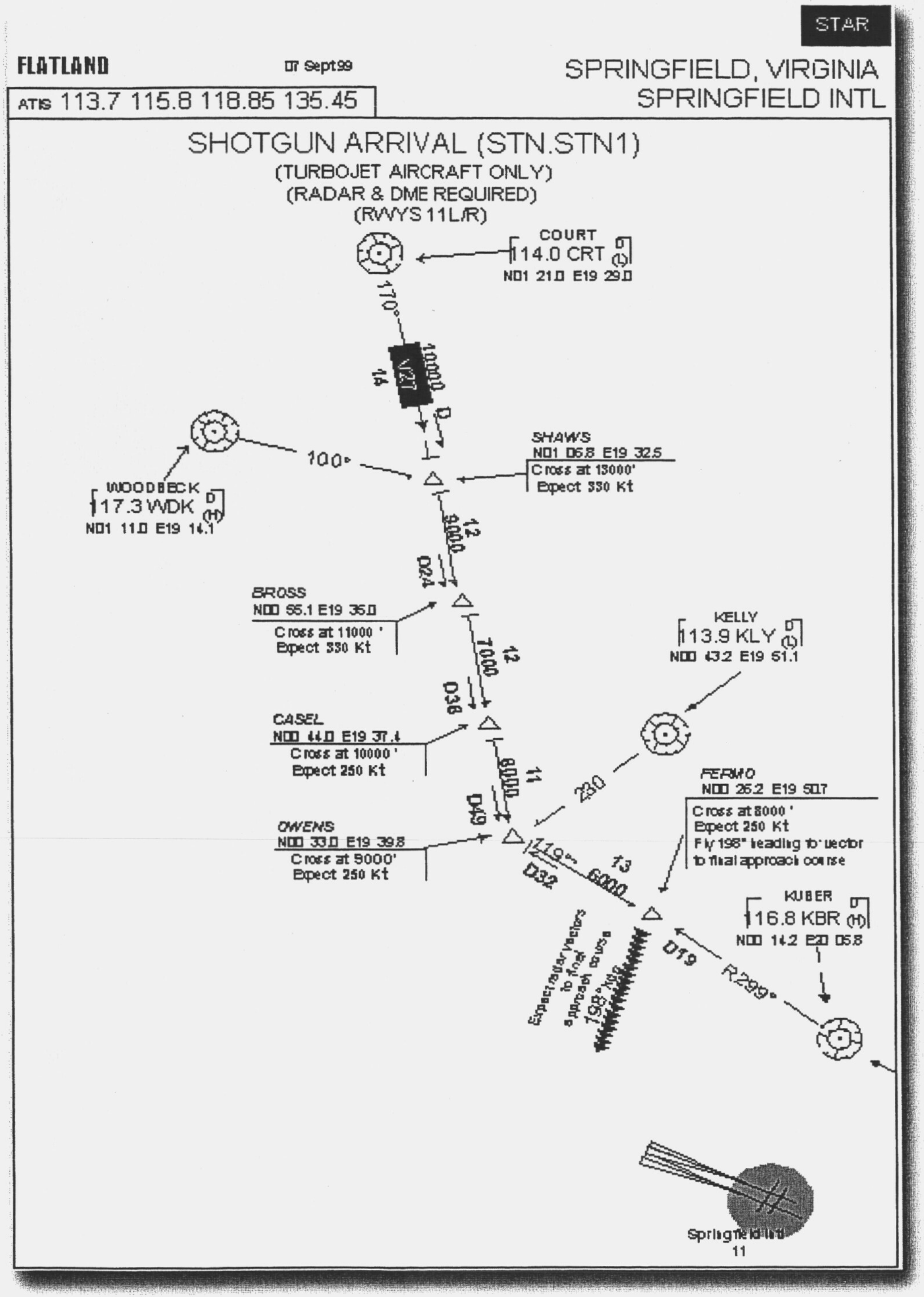


Figure 5 - Sample STAR Chart

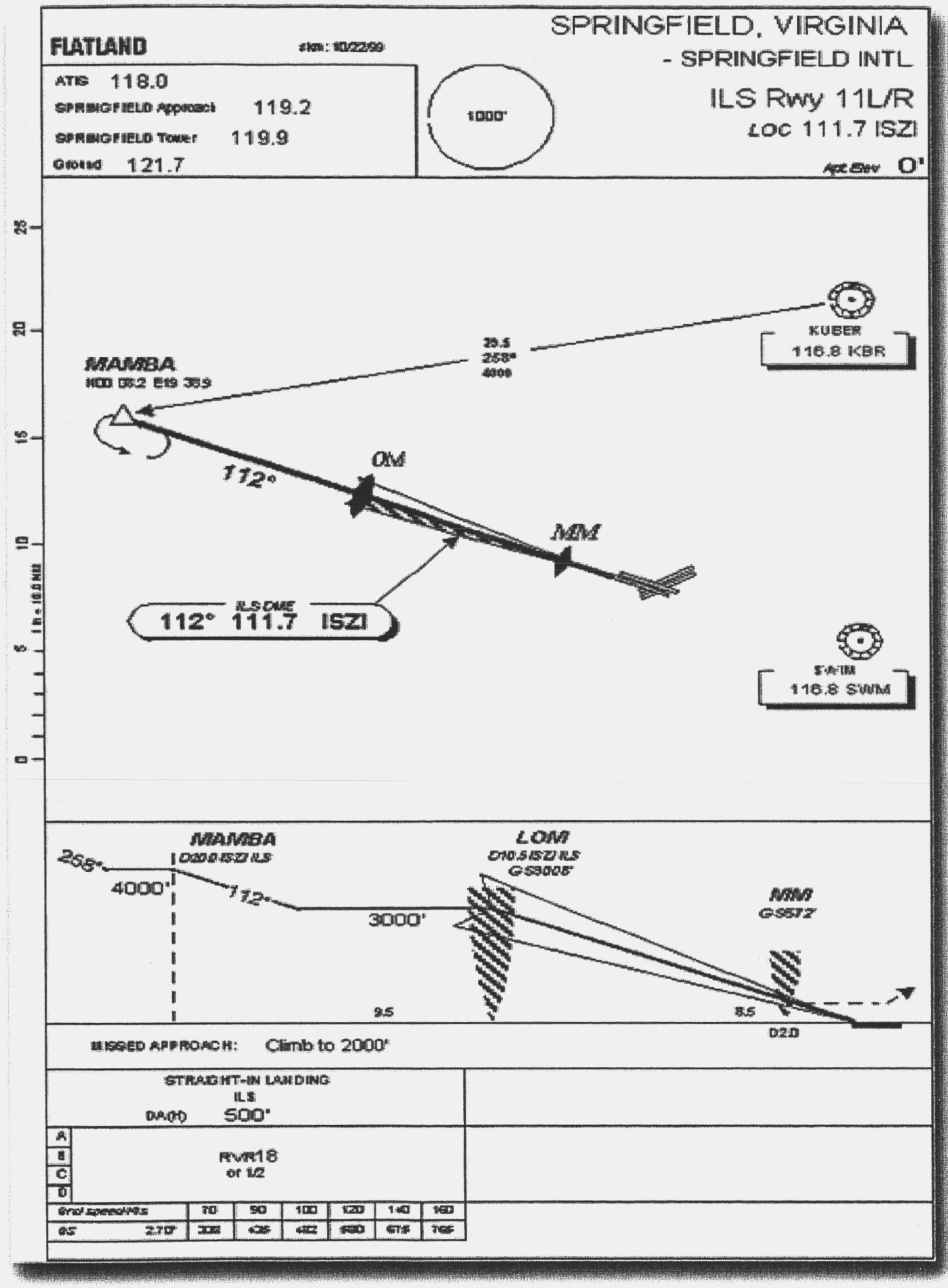


Figure 6 - Sample Approach Plate

Experiment Procedure

The experiment started by getting the informed consent of the participating pilot. This was followed by a briefing about the experiment and the simulator. Prior to the data runs, the pilots were put through training tutorials to acquaint them with the simulator and the experimental setup. This tutorial briefing is supplied in Appendix A. The two tutorials were separated into two phases, one to get acquainted with the various types of automation, and the other to experience a complete scenario. In the first phase of training, the pilot was asked to fly one run using only the MCP. When the pilot was comfortable using the MCP, the first tutorial was restarted and the pilot was exposed to the CDU type of automation and its variants. This phase of training was repeated till the pilot verbally expressed a satisfactory level of proficiency and comfort using all the types of automation. This was followed by the second phase of training, where the pilot was asked to fly a complete scenario using all the automation types to give him a better understanding of what to expect during the data runs. Following the completion of training, the pilots were shown the questionnaires that would follow all experimental runs. Upon completion of both tutorials, the pilots were then given the choice to review any of the previous tutorials or to continue on with the actual experimental runs.

Following the tutorial session, a total of nine scenarios (including the faulty Autoplan scenario) were run for each pilot. For each of the scenarios, the pilot was given a description of the scenario in a briefing sheet and also told what type of automation they would be given. In addition, pilots were also told that all pertinent checklists had been completed and they had only to plan up to the termination point. A first officer was present during all the runs to start the runs, monitor aircraft systems, deploy the flaps and gears as requested and communicate ATC clearances to the test pilot. The first officer played no part in the planning task. In all the runs, the pilots were told the type of automation to use. In the CDU (and its variants) conditions, the pilot was not allowed to use the MCP except to make changes in the altitude window (this was needed since in typical MCP-FMS operation, the aircraft will not climb above or descend below the altitude specified in the MCP altitude window).

Following each scenario, the pilot was given a set of questions pertaining to that scenario. At the conclusion of all the data runs, the pilot was given a brief set of questions pertaining to their background, the experiment as a whole, in-flight replanning and planning tools.

Experiment Participants

A total of sixteen pilots participated in the experiment. Fifteen pilots were from a major airline carrier and one from a major charter service with experience in a major airline service. One pilot was recently retired. All the subjects were male. All the subjects were either captains or first officers.

Total piloting hours ranged from 5000 to 16,000. Eight of the test subjects were captains with experience ranging from 12,000 to 16,000 hours and an average of 12,250 hours of flying experience. The other eight were first officers with experience ranging from 5000 to 10,500 hours and an average of 7400 hours of flying experience. Eleven pilots were initially military trained before becoming civilian pilots and 5 pilots were initially trained in civil aviation. The subjects had flown or were current in a range of glass-cockpit aircraft, including the Boeing 737-800, 737-300NG, 757, 767, and MD-88. Of the 16 pilots, 6 had previous experience with flight planning software of some sort before (other than the FMS), with all six being exposed to ground based planning software and one pilot with experience in ground based (B.A.R.T) and in-flight replanning software (Global Data Systems). All subjects were compensated for their time.

Experiment Apparatus

The experiment was conducted on a fixed-base desktop flight simulator based on the Boeing 747-400. The flight simulator has been developed using the Reconfigurable Flight Simulator (RFS) software (Ippolito and Pritchett, 2000). The simulator runs on two networked desktops PCs. One screen shows the flight instruments, namely, the Primary Flight Display (PFD), Electronic Horizontal Situation Indicator (EHSI) (also known as the Navigation Display [ND]), and controls for the flaps and gears. The second screen displays the Mode Control Panel (MCP), the Control Display Unit (CDU) and navigation display controls (ND controls). Both the desktops PCs were equipped with a mouse as an input device. The setup was distributed over four flat panel LCD screens with two screens - one displaying the PFD, EHSI and flaps and gears, and the other displaying the CDU, MCP and ND controls - for the captain and two screens showing the same displays for the first officer. Figure 7 shows the experiment setup.

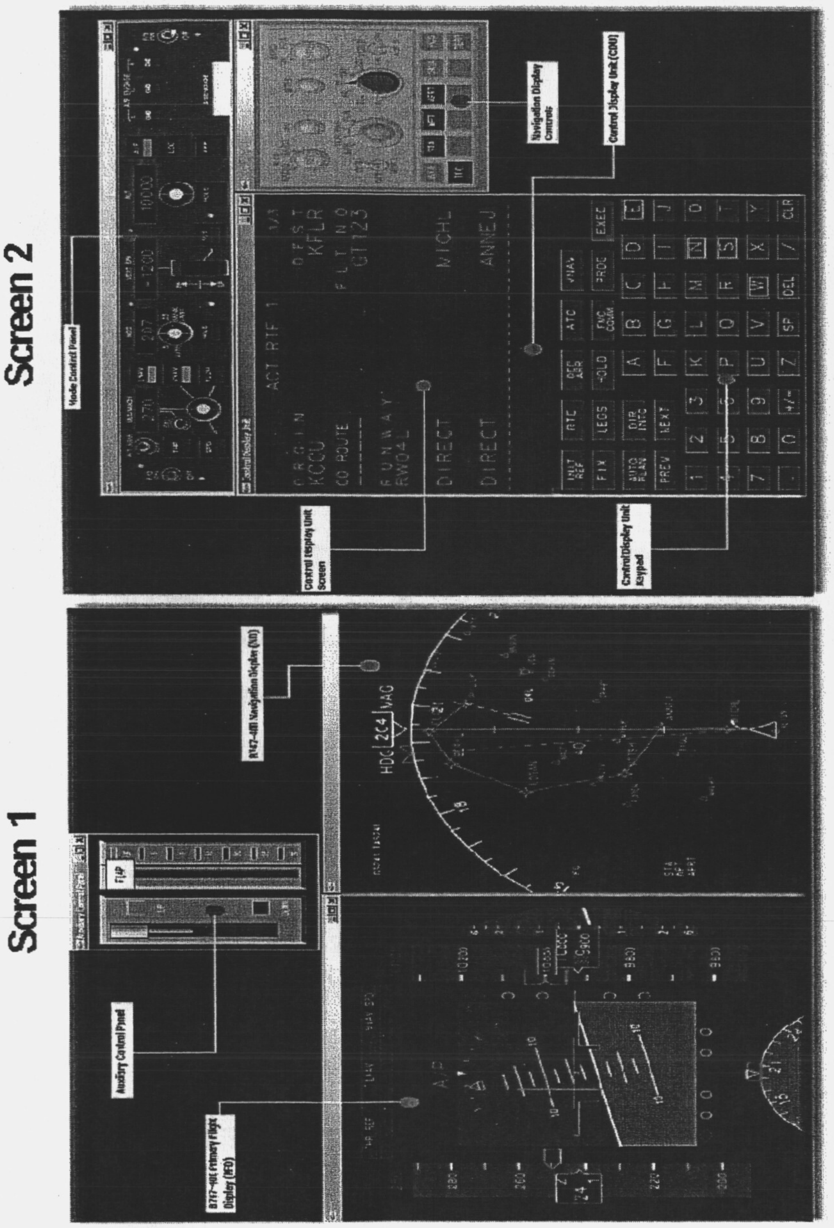


Figure 7 - Experiment Setup for Each Pilot

The flight instruments included the primary flight display (PFD), electronic horizontal situation indicator (EHSI), the Mode Control Panel (MCP) and Control Display Unit (CDU), all of which are based on the Boeing 747-400 glass cockpit.

The PFD (Figure 8) shows the current aircraft state such as the current airspeed and altitude. At the top center of the PFD are the Flight Mode Annunciators (FMAs) which display which mode of flight the autopilot is in. The magenta figures above the altitude and speed tapes show the MCP target altitude and target speed respectively. The vertical speed indicator beside the altitude tape shows the rate of climb or descent. The two magenta bars in the middle of the display are the Flight Directors (F/D) which show the pitch and roll of the aircraft. The arrow indicator at the top of the calibrated scale on the artificial horizon indicates the bank angle.

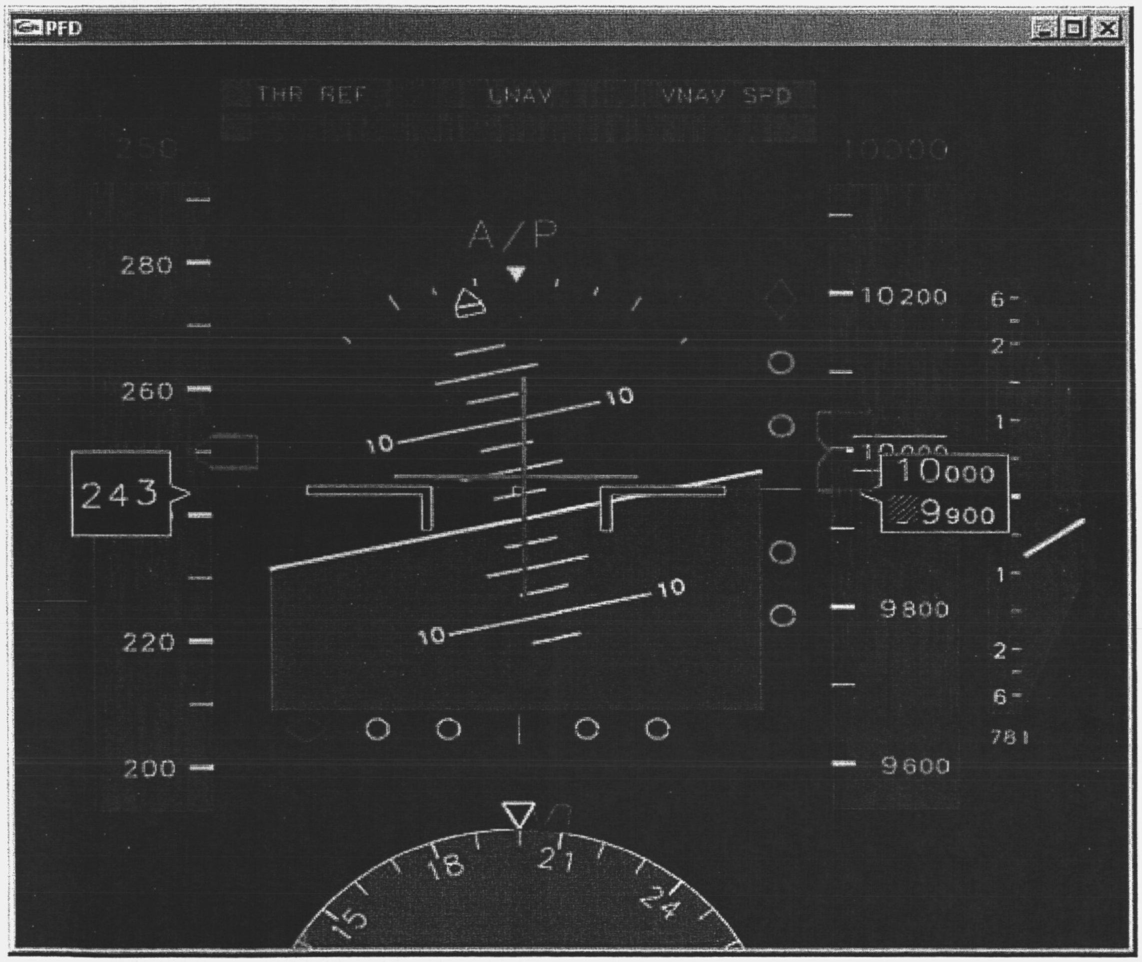


Figure 8 - Primary Flight Display (PFD)

The EHSI (Figure 9) used in this experiment is based on that used in the B747-400. The EHSI is comprised mainly of a track up moving map display. The display shows the current flight path as a solid magenta line. Any lateral modification to the current active flight path is shown by a white stippled line. The current position of the aircraft is shown as a solid white triangle. The green arc shows the point where the aircraft will reach its MCP target altitude. The map also shows the various navigation aids (with their identifiers) in the vicinity of the aircraft in blue. The destination runway is shown in white with its 3-letter identifier, with the approach line extending 14 miles. The Autoplan shows up on the EHSI as a stippled orange line which turns solid magenta when executed.

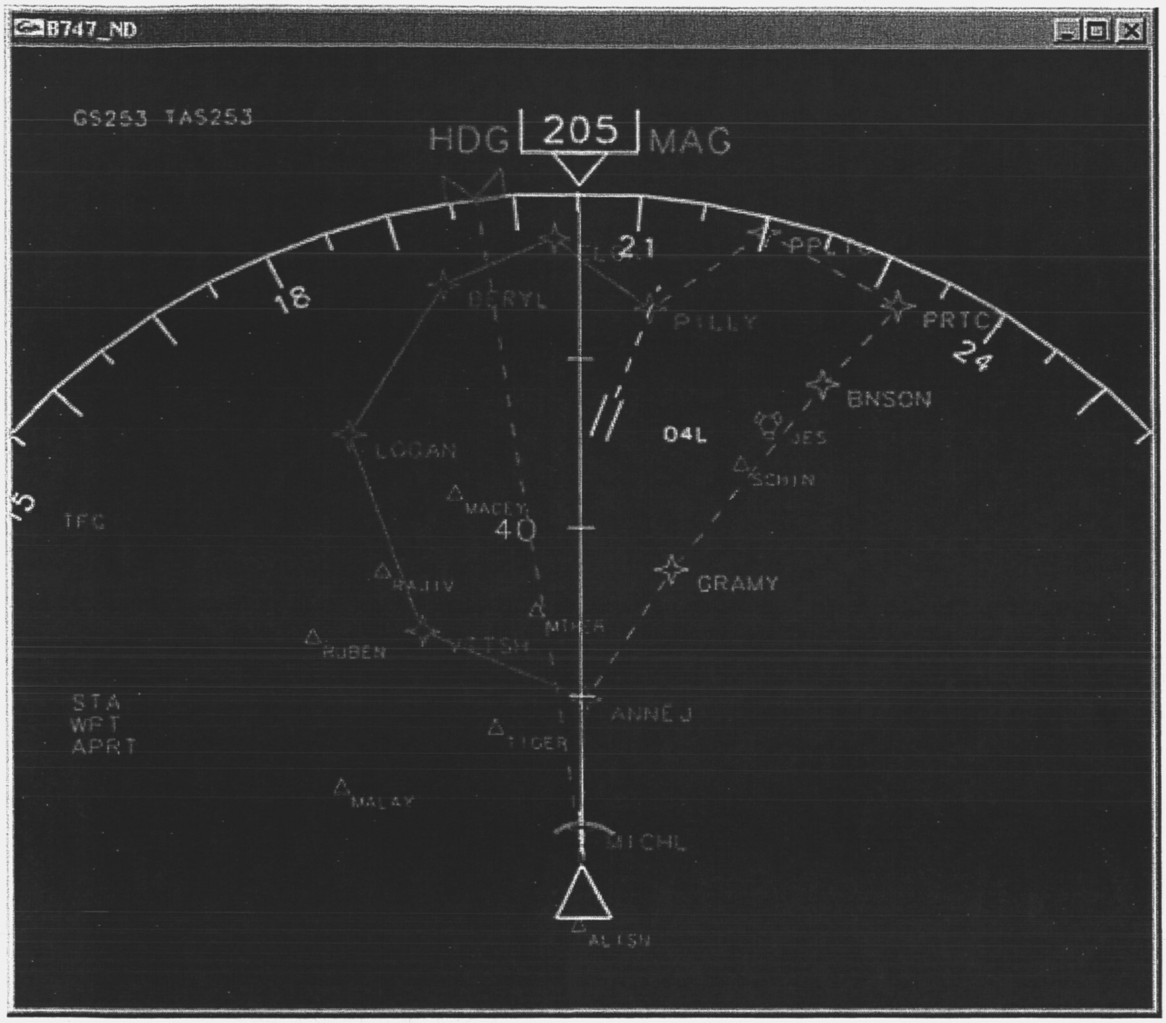


Figure 9 - Electronic Horizontal Situation Indicator (EHSI)

The MCP is an autoflight system through which the pilot can change heading, altitude, speed and rate of descent. The flight mode (i.e., HDG, FLCH, VS, ALT, LNAV, VNAV, and SPD) selected in the MCP is displayed on the FMA on the PFD. The MCP used in this experiment (Figure 10) is modeled on the B747-400 MCP, and the pilot used a mouse as an input device to enter values into the MCP. The target values for speed, heading, vertical speed and altitude could be entered by the pilots by clicking on the dials below the display window. For example, to change heading, clicking on the right half of the circular dial will increase the heading angle and clicking on the left half will decrease the heading angle, and similarly for the Indicated Air Speed (IAS) and altitude. The vertical speed (V/S) is usually controlled by a roller dial which in this MCP is the pink and indigo dial just below the V/S target window.

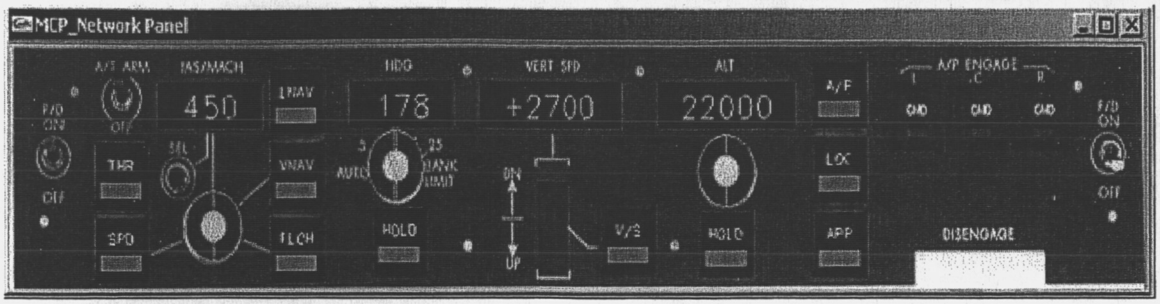


Figure 10 – RFS Mode Control Panel

The CDU is an autoflight system which, among other things, pilots use to plan/replan flight routes. This experiment used a graphical interface CDU (Figure 11) modeled on the B747-400 CDU, where the pilot used a mouse as an input device to enter data into the CDU. For this experiment, the pilot had only the RTE and LEGS pages available to them. Pilots could enter data into the scratchpad and insert it wherever desired.

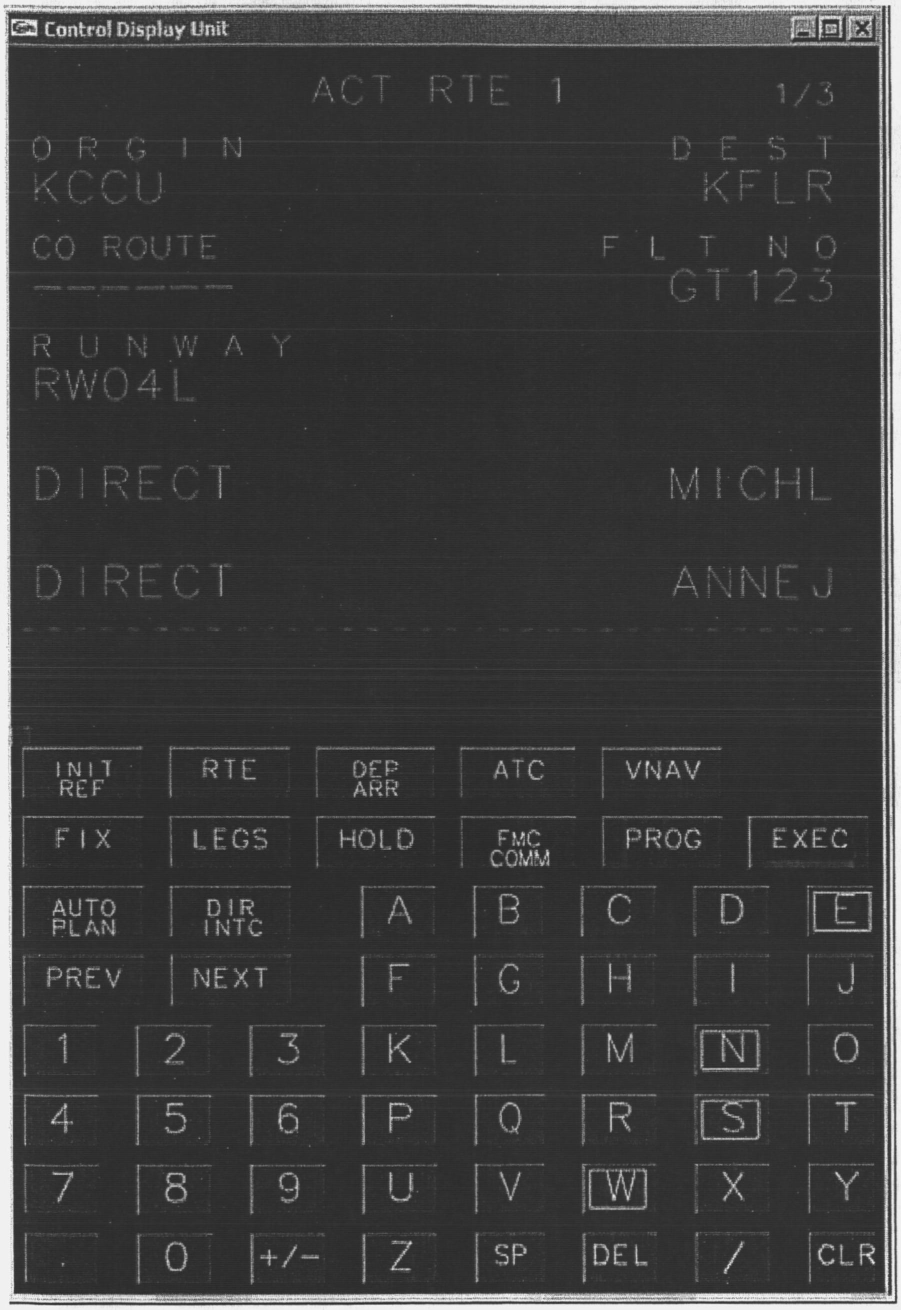


Figure 11 - RFS Control Display Unit

Independent Variables

Two scenario types were tested, namely, emergency and non-nominal situations. Both of these required the pilot to perform tactical planning.

Non-Nominal Scenarios: These are situations where there is no unusual urgency to land the airplane. These are not very important in terms of the time taken to land. All of these cases can be resolved with a simple detour from the original flight plan. The non-nominal scenarios used in this experiment were:

- *Runway Closure:* Required the pilot to reroute to a nearby alternative.
- *Runway Change:* Required the pilot to change the destination runway.
- *Weather Disturbance:* Required a pilot to navigate around a weather disturbance i.e., a storm cell.
- *Opening up/closing of restricted airspace:* Required a pilot to navigate around restricted airspace.

Common to these scenarios is the fact that they envision landing in the order of tens of minutes, i.e., immediate landing is not an overwhelming concern. Other factors such as aircraft stability, fuel economy, standard operating procedures, etc. are important factors when deciding on the rerouting. None of these conditions alter the performance of the aircraft in any way and fuel was not a concern.

Emergency Scenarios: These are situations where there is an urgency to land the aircraft as soon as possible. Thus, in the event of emergencies, the pilots are given a free hand in deciding the route to be taken which may involve violating any altitude and speed constraints or procedures. Emergency situations can have a number of causes. The emergency scenarios used in this experiment were:

- *Cargo Fire:* This is an emergency wherein a fire in the cargo hold had just been extinguished at the start of the run. The extent of damage was not known and the pilot was required to land the aircraft as soon as possible.
- *Medical Emergency:* This emergency required the pilot to replan, reroute and land as soon as possible.
- *Fuel Filter Emergency:* This is an emergency wherein the fuel filter can get blocked by debris thereby inhibiting the intake of fuel into the engines. Landing immediately is imperative.
- *Loss of Hydraulic Pressure in One of the Hydraulic Systems:* This is an emergency wherein the EICAS shows a loss of hydraulic pressure in one of the hydraulic systems. Landing immediately is imperative.

All of the emergencies were predicted to be of equal severity. However, they are similar in that the replanning process still has to be executed and the new route implemented, and they do not alter the performance of the aircraft in any way. Emergency scenarios differ from non-nominal scenarios in that they envision the time to landing to be less, i.e., on the order of a few minutes.

In each run, the pilot was asked to use a particular type of automation. Specifically, the four types of automation tested are detailed were:

- *Mode Control Panel (MCP):* Pilots were only allowed to command the following autoflight modes through the MCP using heading select and heading hold (HDG), vertical speed (V/S), altitude hold (ALT), flight level change (FLCH) and speed (SPD).

- *Control Display Unit (CDU):* In this condition, pilots were asked to use a conventional CDU based on a Honeywell 747-400 CDU. Only pages that assist in planning (RTE and LEGS) were made available to them.
- *Control Display Unit + (CDU+):* With this type of automation, pilots had the CDU available to them as in the previous case. This automation had an added functionality called the Autoplan. This is a computer generated flight path that can assist pilots in planning. Pilots could access these plans whenever they like and use it as the active route, or plan so that their route can intersect parts of the Autoplan, or disregard it totally. The Autoplan feature does not exist in current cockpits.
- *Control Display Unit ++ (CDU++):* This automation works the same as the CDU+ with the difference that, when the simulation run starts, the Autoplan is implemented as the active flight route. Pilots have the option of overriding this plan or modifying as in the previous automation. In the CDU+ and CDU++, the Autoplan was designed to be the best plan for the given scenario type. For example, in the emergency scenarios, the Autoplan was designed to get the aircraft down as soon as possible, keeping in mind standard airspace regulations and following/intersecting standard airways as depicted in the charts. In the non-nominal scenarios, the Autoplan placed stress on other factors such as negotiating the cause of re-route and minimizing the distance flown.

Experiment Design

The experiment was divided into two parts run sequentially in one session. The first experiment tested all eight combinations of automation and scenario types. In the second experiment, pilots were asked to fly only one run, the ninth run, using the CDU++ type of automation only. The experiment condition in this run was based on the same automation-scenario combination for all the pilots. The second experiment was included in the tests to explore the effect of an erroneous automatically generated plan on pilot performance. This faulty Autoplan scenario followed completion of the primary scenarios.

The first experiment consisted of a 4x2 test matrix as shown in Table 5, and was made up of a combination two independent factors, type of automation and scenario type. The test matrix was arrived at by first blocking by type of automation. Then, within each block of type of automation, the two scenarios types were run in random order. The order of the automation block was based on a fully balanced Latin squared design to mitigate order effects. Specific scenarios were assigned randomly and care taken that the same number of pilots flew the same scenario with the same automation.

Table 5. Experiment Test Matrix

Type of Automation	Scenario Type: Non Nominal	Scenario Type: Emergency	
MCP	16 Pilots x 1 run	16 Pilots x 1 run	Expt. #1
CDU	16 Pilots x 1 run	16 Pilots x 1 run	Expt. #1
CDU+	16 Pilots x 1 run	16 Pilots x 1 run	Expt. #1
CDU++	16 Pilots x 1 run	16 Pilots x 1 run	Expt. #1
Faulty Autoplan	All runs using CDU++		Expt. #2

Expt. #1:

- 8 runs per pilot
  - 4 runs non-nominal
  - 4 runs emergency
- Run order blocked by automation and balanced using Latin Square Design

Expt. #2:

- 1 run per pilot
  - ½ pilots had non-nominal scenarios
  - ½ pilots had emergency scenarios

The second experiment consisted of a 1x2 matrix, and was made up of a combination of one type of automation, the CDU++, and the two scenario types. It consisted of only one run per pilot, and used a between subjects design, where the scenario types were randomly assigned. This experiment used the faulty Autoplan scenario where an error in the automation provided the pilot with an inappropriate Autoplan. The plan lacked context sensitivity to the situation and thus did not provide the best plan for the current situation, i.e., in the non nominal flight condition the Autoplan generated an overly aggressive route fit only for emergencies and, in the emergency flight condition, provided a gently paced route that increased time of flight beyond what the emergency called for.

Dependent Measures

Three types of data were collected:

1. The graphical interface of the CDU recorded important events in the flight replanning task. The final mouse click triggering an event was recorded as an identifiable action. These events included switching between RTE and LEGS pages, making changes to an existing RTE page or a LEGS page, going through the route programmed in by clicking the PREV and NEXT buttons (in the case of the CDU-based autoflight conditions), making altitude, speed and heading changes (in the case of the MCP), creating/deleting a fix/waypoint from the flight plan, changing altitude and speed parameters of existing waypoints, resolving route discontinuities, looking at alternative routes, activating an inactive route, and executing a change in the flight plan.

2. Aircraft state data, including airspeed, current heading and current altitude, was logged every second by the simulator.
3. At the end of each run and at the end of the experiment, pilots were asked to answer a questionnaire. The end of run questionnaire included questions about the factors considered during planning, the strategies used, and effectiveness of the autoflight system used, a rating of the ease of planning using that autoflight system compared with currently available type of automation they would have used and a NASA TLX workload rating sheet. The end of experiment questionnaire included questions on pilot background, in-flight replanning in the two scenario types, flight replanning systems and tools, performance of the Autoplan and the NASA TLX pair-wise comparisons of sources of load. The complete end of run questionnaire and end of experiment questionnaire are given in Appendix B.

Analysis of the aircraft state data, event logs during the flight, performance logs during the flight, measures such as duration and the length of the run, modifications to the Autoplan and pilots' responses to the questionnaires included:

- Ability to diagnose/recognize errors in automation;
- Pilot dependency on automation;
- Time and distance saved for that run compared with the original plan;
- Pilot's choice of route implemented and the apparent reasons behind the choice. This could indicate the correlation (if any) between type of automation, type of scenario and in-flight replanning behavior;
- Deviation from the existing preprogrammed route to indicate the amount of time saved compared with the time taken if the original path was followed;
- Time taken to start modifying the existing plan or entering a new plan;
- Regularity with which they tend to update the plan versus leaving it once it has been created;
- Apparent strategies and factors considered during planning;
- Pilot preferences of certain autoflight systems for in-flight replanning tasks;
- Comparison of ease of planning using different autoflight systems;
- Performance of the Autoplan; and
- Workload assessment of the replanning task.

The data analysis was divided into three categories: pilot performance, pilot planning behavior, and workload assessment.

Performance was measured in the first experiment by time to landing and distance to landing. In the ninth run, pilot performance was also measured by whether pilots recognized the Autoplan was faulty.

Planning behavior can be manifested in a number of ways. Apparent strategies were analyzed for planning with the MCP and the CDU such as establishing one-dimension of path first followed by the other. For example, some pilots may prefer to plan for the lateral path first and then the vertical path. Others may plan for the vertical path first and then the lateral path so that they don't need to descend at a high rate, yet others may do it as a series of heading changes followed by descents. In some cases, some pilots may start planning immediately and bring out a rough plan and then keep refining that plan over time, whereas others may take some more time and come up with an almost concrete plan which requires few adjustments.

A workload assessment based on pilot responses to the NASA TLX Workload sheet was also performed to examine which source of workload was felt the most during the scenarios.

Results

In total, 144 runs were performed: 128 under the first experiment comparing the different autoflight systems and 16 runs for the second experiment's faulty Autoplan case. The 16 faulty Autoplan runs will be discussed separately from the regular 128 runs.

Unless otherwise specified, the data obtained were analyzed for type of automation, scenario type, specific scenario and run order effects by fitting to a general linear model. If the residuals of the fit met the requirements for Analysis of Variance (ANOVA), an ANOVA was conducted. The type of automation, scenario type and specific scenario were analyzed as fixed effects. Pilots, however, were analyzed as a random factor, allowing generalization of the observations to a major portion of the pilot population. In addition, interactions between the factors were examined.

Where significant results were found for one or more of the factors, a one-way (ANOVA), along with a pair-wise comparison using a 95% confidence level Tukey test, was performed strictly on those factors to confirm the results. A non parametric Kruskal-Wallis test was also performed to test the null hypothesis that there are no differences among the factors.

Pilot Performance

The primary measures of pilot performance in the first experiment were the distance flown and the duration of the run. In emergency situations, such as a medical emergency or cargo fire, these measures directly reflect the safety of the aircraft. In non nominal situations, such as weather or airport closures, these measures reflect airline operation considerations such as flight time, flight schedules and fuel burn. As can be seen in Figures 12 and 13, in both scenario types, the type of automation used was not a significant factor. No significant order effects were seen on these measures either. The scenario type and the specific scenario, however, did show a significant effect on these measures. To confirm the scenario and scenario type effects, an ANOVA was performed, showing a significant scenario effect ($F = 33.92$, $p < 0.001$) and an effect from the scenario type ($F = 69.46$, $p < 0.001$).

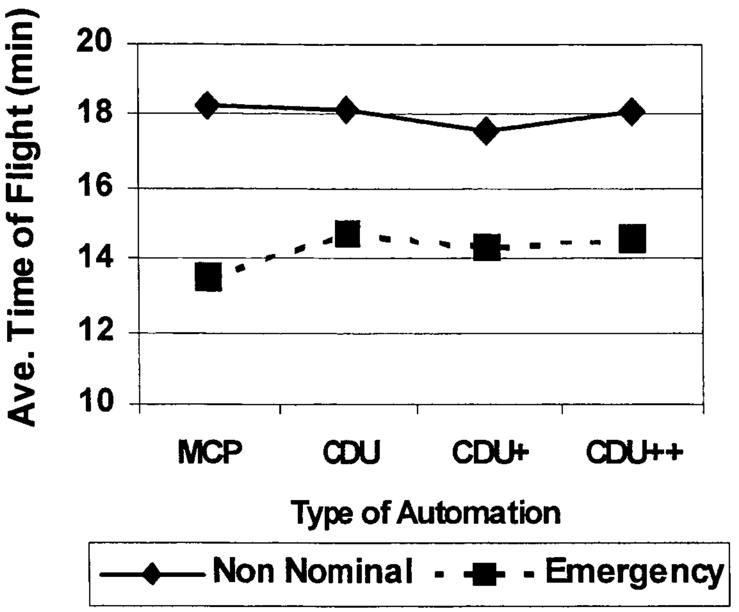


Figure 12 - Average Time of Flight

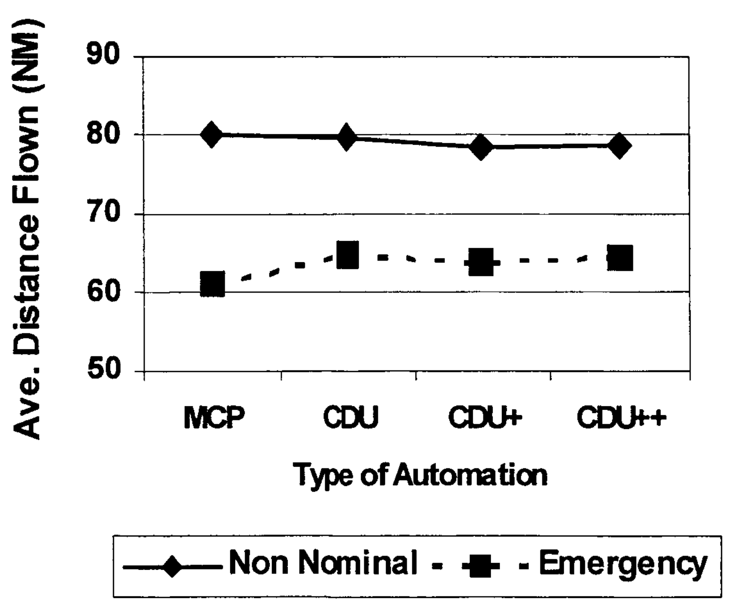


Figure 13 - Average Distance Flown

Since the type of automation did not have any significant effect on the performance measures of time and distance, these measures were also looked at by specific scenario across all types of automation. As seen in Figures 14 and 15, in the non nominal scenarios, the average time of flight and distance flown were distinctly higher for the first two scenarios: weather disturbance and restricted airspace. A certain degree of variability in flight path was seen. In the emergency scenarios, time of flight and distance flown was higher for the second two: the hydraulic systems failure and fuel filter emergencies.

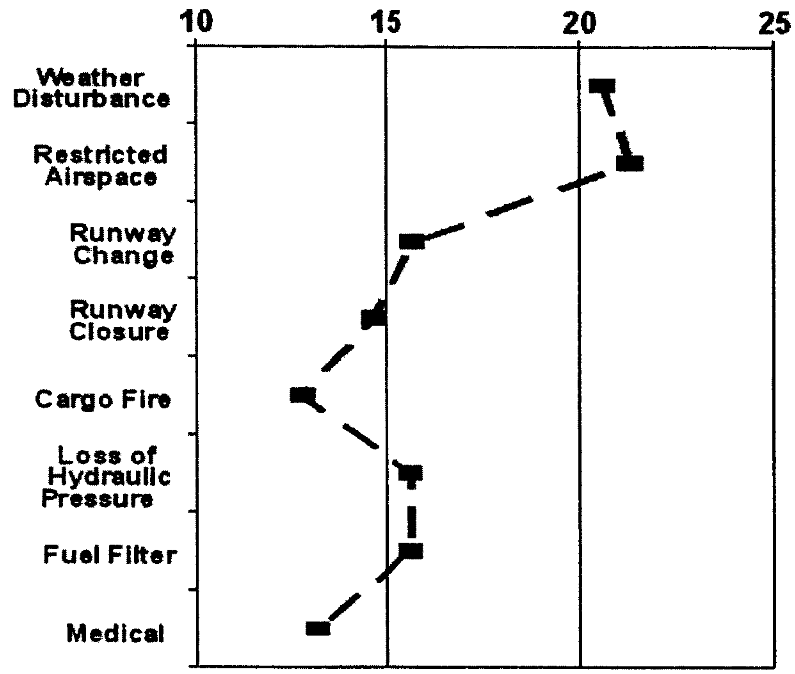


Figure 14 - Average Time of Flight per Scenario

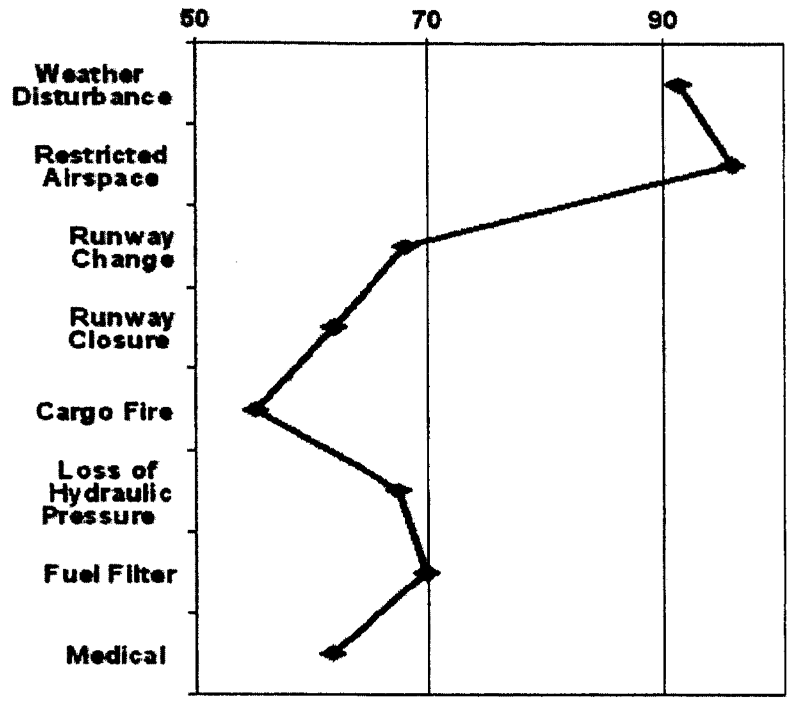


Figure 15 - Average Distance Flown per Scenario

Although the scenarios were intended to have similar travel times, to account for any intended differences between scenarios, another measure was the deviation in time of flight and distance flown from the baseline plans for each scenario. The baseline plans used were the original routes in the CDU at the start of the run. As can be seen in Figures 16 and 17, the main effects here were also the scenario type ($F = 66.43$, $p < 0.001$) and the scenario ($F = 14.72$, $p < 0.001$). Additionally, no run order effects were seen with the time of flight measure, but the distance flown showed significant run order effects ($F = 9.66$, $p=0.003$) (Figure 18).

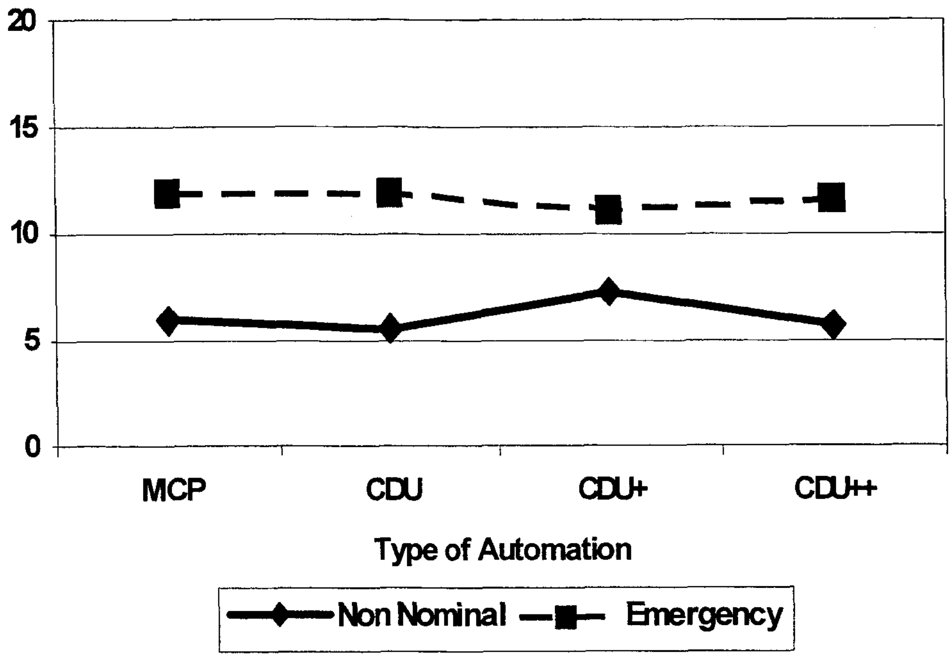


Figure 16 - Average Deviation in Time of Flight

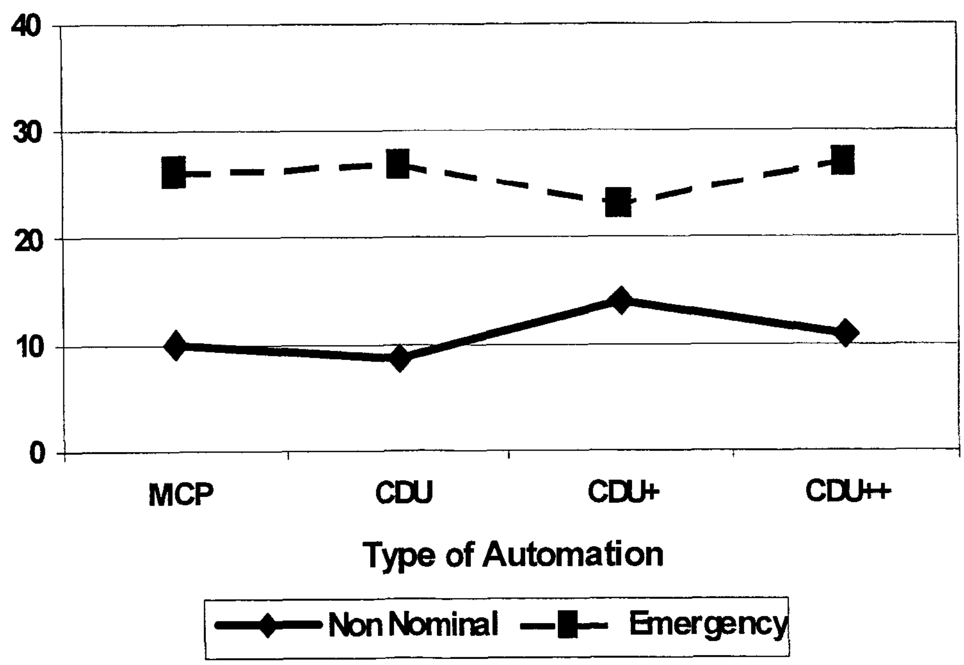


Figure 17 - Average Deviation in Distance Flown

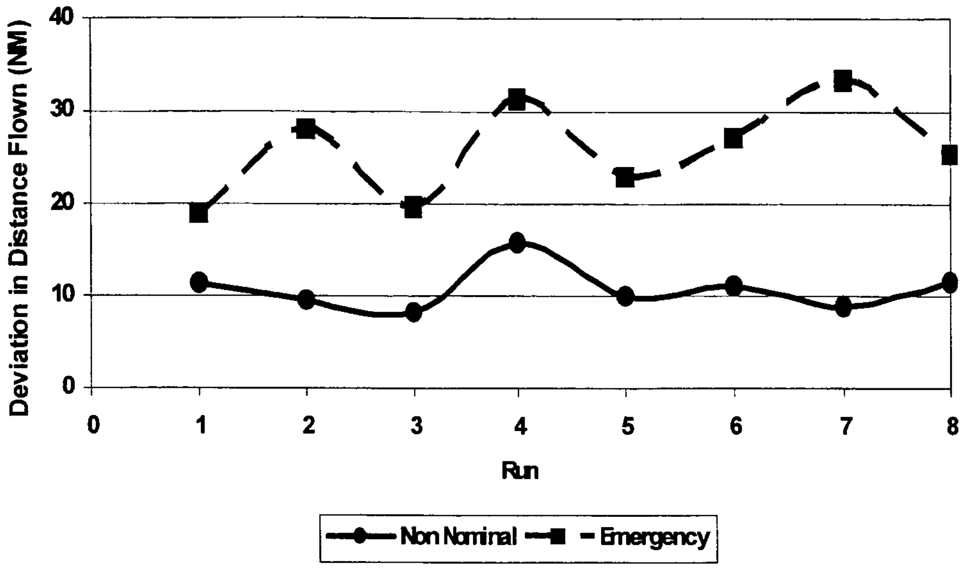


Figure 18 - Run Order Effect on Deviation in Distance Flown

Another measure pertaining to pilot planning performance were the speed violations. According to Federal Aviation Administration (FAA) regulations, aircraft flying below 10000 feet must remain at a speed of 250 knots or below, except when given discretion by a controller or in an emergency. The scenario, scenario type and run order showed significant effects on this measure as did the pilot-scenario interaction. However, all these effects failed normality tests and an ANOVA could not be conducted. Order effects were also seen but also failed subsequent normality tests. However, the non-parametric Kruskal-Wallis test showed that the medical emergency (emergency) had the highest number of speed violations and the weather disturbance (non-nominal) had the lowest number of speed violations.

Pilot Planning Behavior

A number of measures examined pilot planning behavior. First the flight paths were looked at for any trends in planning behavior. As an example, the flight paths are shown in Figure 19 for the medical emergency scenario. During the experiment it was observed that the type of automation and scenario had an effect on pilots' course of action. For this reason, to describe the effect of the automation on planning behavior, the measures were also analyzed by the type of automation. The specific scenario effect was also considered to explain specific behaviors.

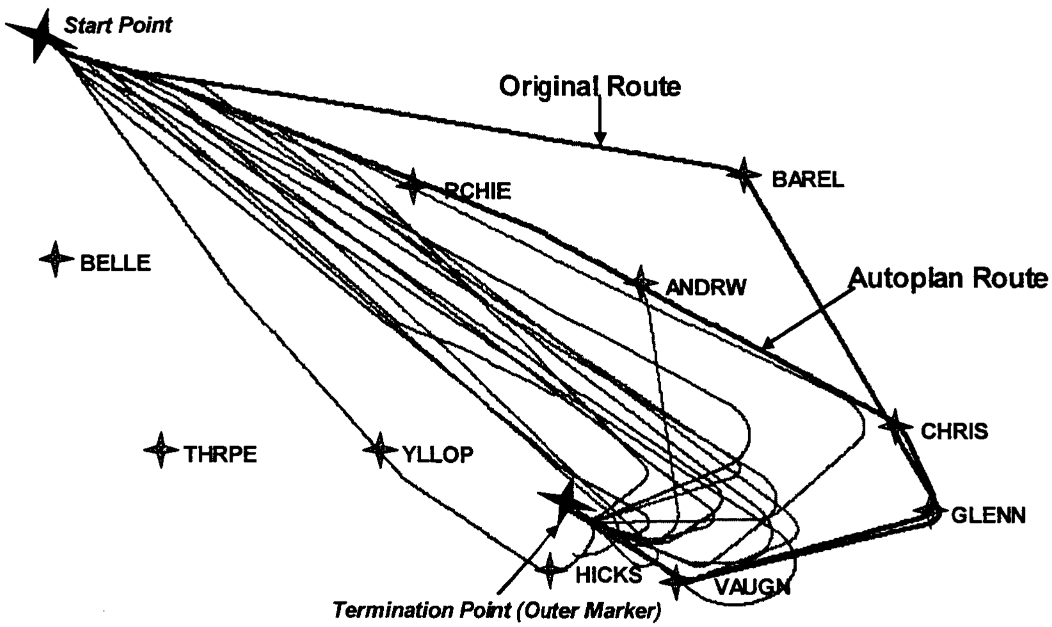


Figure 19 - Flight Paths for the Medical Emergency Scenario

General Observations on Planning Behavior

In general, in each scenario type, the primary objectives were to minimize distance to go and to create an expeditious route to the approach. The timing and ordering of fixes did not show any specific pattern by which a pilot tended to plan. The usage of the type of automation also showed very specific personal choice traits. For example, 10 of 16 pilots, with all types of automation, immediately increased speed and kept a high altitude to get abeam of the outer marker as fast possible. All pilots except one created an along track waypoint ahead of the waypoint being flown direct to, at which point they started reducing speed. In 58.3% of the runs where pilots could create along track waypoints (CDU, CDU+ and CDU++ types of automation = 96 runs), this point was abeam of the outer marker, which would give them a much smoother turn onto the final approach leg. When using the MCP, this point was visually marked out (as verbally reported by pilots during the experiment) and then HDG SEL was used to turn onto final. When using the CDU and its variants, all pilots used the only the LEGS page during planning as this provided the necessary information of heading, distance, and speed and altitude constraints at waypoints. In general, 14 pilots agreed with the routing the Autoplan provided; however, they did not agree with the speed and altitude profile in the Autoplan and proceeded to make subsequent changes. Most pilots used the Autoplan to orient themselves in the desired direction and then modified the waypoints to create a more direct route to the runway.

In terms of dimensional planning, in all cases, pilots first got themselves oriented in the desired direction. This was then followed by a 'cleaning up' of the route, where some waypoints were deleted or added to provide a more direct route. This was then followed by a series of speed and altitude changes until the termination point. Speed and altitude changes did not follow any specific pattern.

Some interesting observations in usage of the type of automation were made during the experiment. Some of the pilots, in order to reduce workload, would simply 'trick' the CDU into behaving like an MCP. For example, in a long stretch, the pilot would put in a very low altitude constraint at the active waypoint, which in turn would provide a high rate of descent, and then change the altitude constraints back to specified limits when at a suitable distance from the waypoint. Only two pilots resorted to this technique as they did remember that they would not be allowed to use the MCP when using the CDU or its variants.

Another technique commonly used by the pilots was the DIRECT-TO function. In some cases, instead of creating a waypoint abeam or a little ahead of the marker, pilots would wait till the aircraft was abeam or a little ahead of the marker and then initiate a DIRECT-TO to the outer marker after accounting for the distance required for a turn. This proved extremely effective, and had a result similar to that of creating a waypoint, albeit the turn required was sharper. This behavior was exhibited in five runs spread among two pilots. In all scenarios, the pilots were cleared to the glideslope altitude. Thus, when using the CDU and its variants, in most cases, pilots would enter the clearance altitude into the altitude window in the MCP and then adjust the vertical profile by altitude changes in the CDU LEGS pages. This was done to eliminate the altitude intervention by the MCP.

Pilot Planning Across Automation Types

In the MCP cases, whenever a new speed or altitude target was entered and kept constant for at least fifteen seconds, it was counted as a speed or altitude change. For the CDU cases, the change in a future speed or altitude or both was identified as an event and logged in the simulator. It was observed that the average number of speed and altitude changes when using the MCP was distinctly lower than for the other types of automation which is evidenced by Figure 20 and Figure but did not show any statistical significance. This suggests that with the MCP, pilots did not have the hindrance of forcibly changing speed and altitude constraints at waypoints as was required with the other TOAs. Among all the types of automation, the CDU+ was found to have the highest average number of speed and altitude changes. This suggests that these measures are more a function of pilot choice than scenario or automation effects.

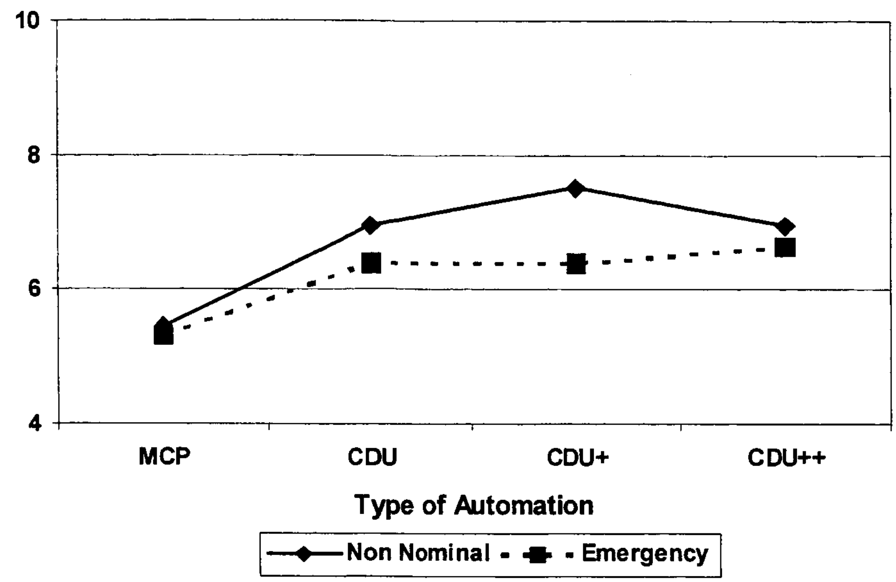


Figure 20 - Average Number of Speed Changes

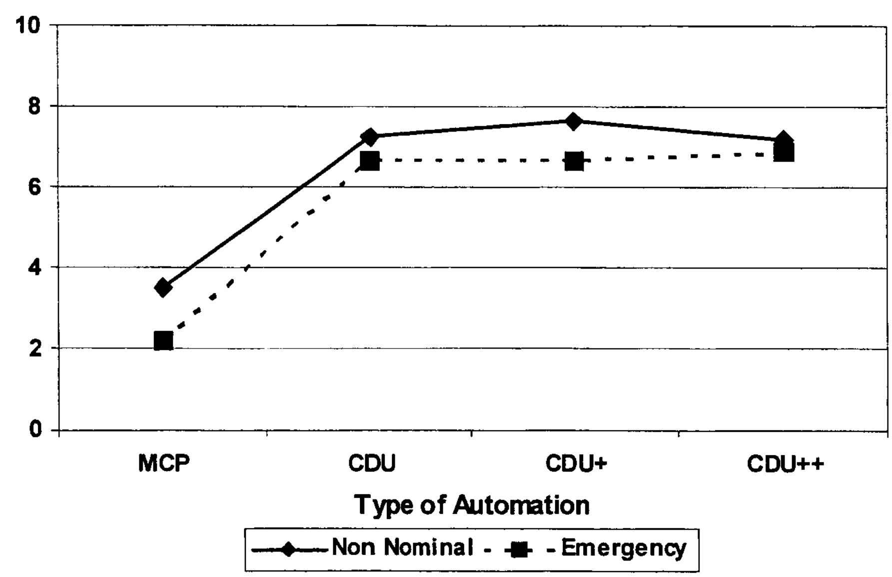


Figure 21 - Average Number of Altitude Changes

Two more measures examined pilot behavior with the CDU (and its variants). These were the time taken to the first modification and the time taken for the first execution of a change to the route in the CDU from the start of the run. The time taken to first modification was defined as the time difference between the start of the run and the first instance when the page status (either the RTE or LEGS page) changes from active (ACT) to modified (MOD), and the time taken for the first execution was defined as the time difference between the start of the run and the first instance of the

EXEC button being pressed to confirm an action. These measures were indicative of the time the pilot takes to start planning and implement a change to the plan. A combination of these two measures showed that on an average, pilots took a shorter time to start re-planning using the CDU+ type of automation than with CDU or CDU++.

In addition to the above, apparent strategies in planning were also examined. A general pattern that did emerge was that pilots oriented themselves in the desired direction first (mostly direct to a point abeam the marker) by either using HDG SEL in MCP cases or initiating a DIRECT-TO in the CDU (and its variants) cases. This was followed by vertical profile management via speed and altitude changes to get to that point, followed by a turn to base leg to line up for approach fully configured. Figures 22 and 23 show the real paths and planning pattern for a non-nominal scenario (weather disturbance) and an emergency scenario (loss of hydraulic pressure) respectively.

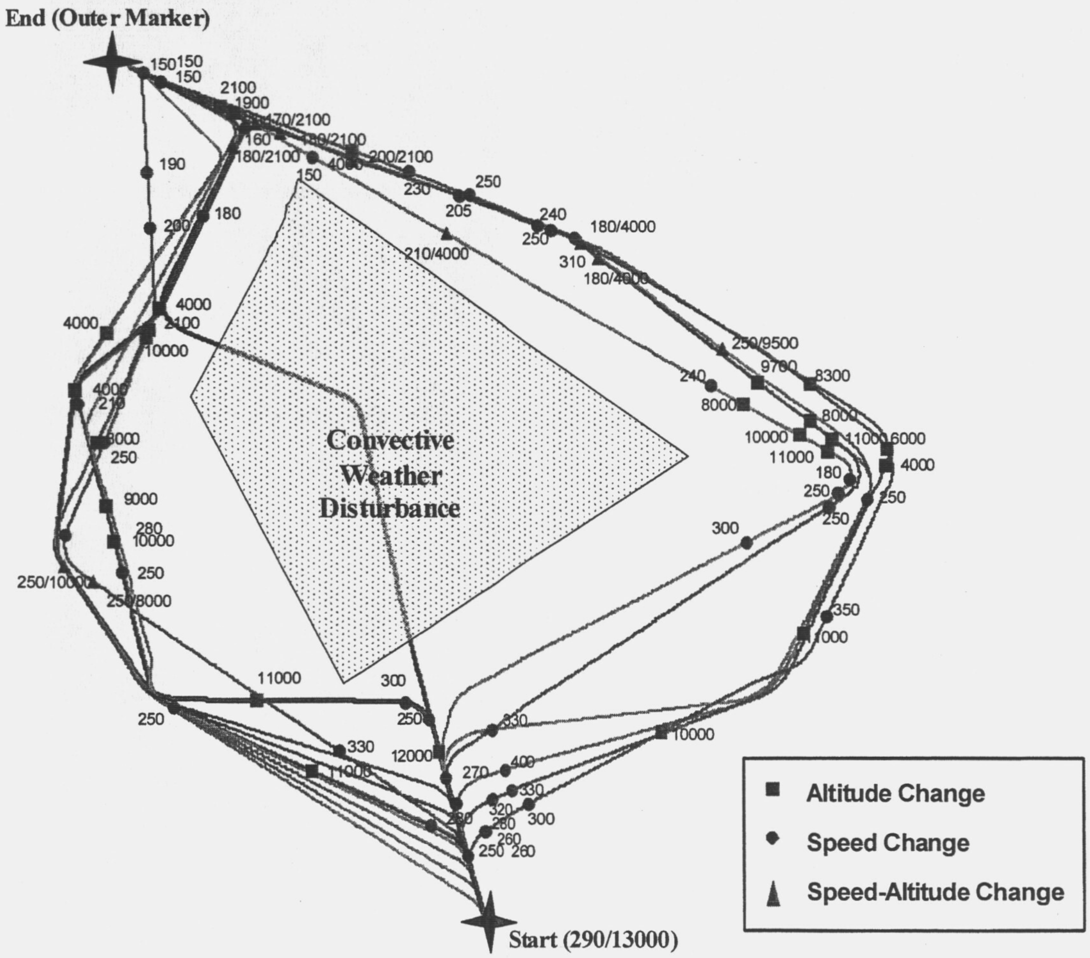


Figure 22 – Flight Paths and Altitude and Speed Changes in the Weather Disturbance Scenario

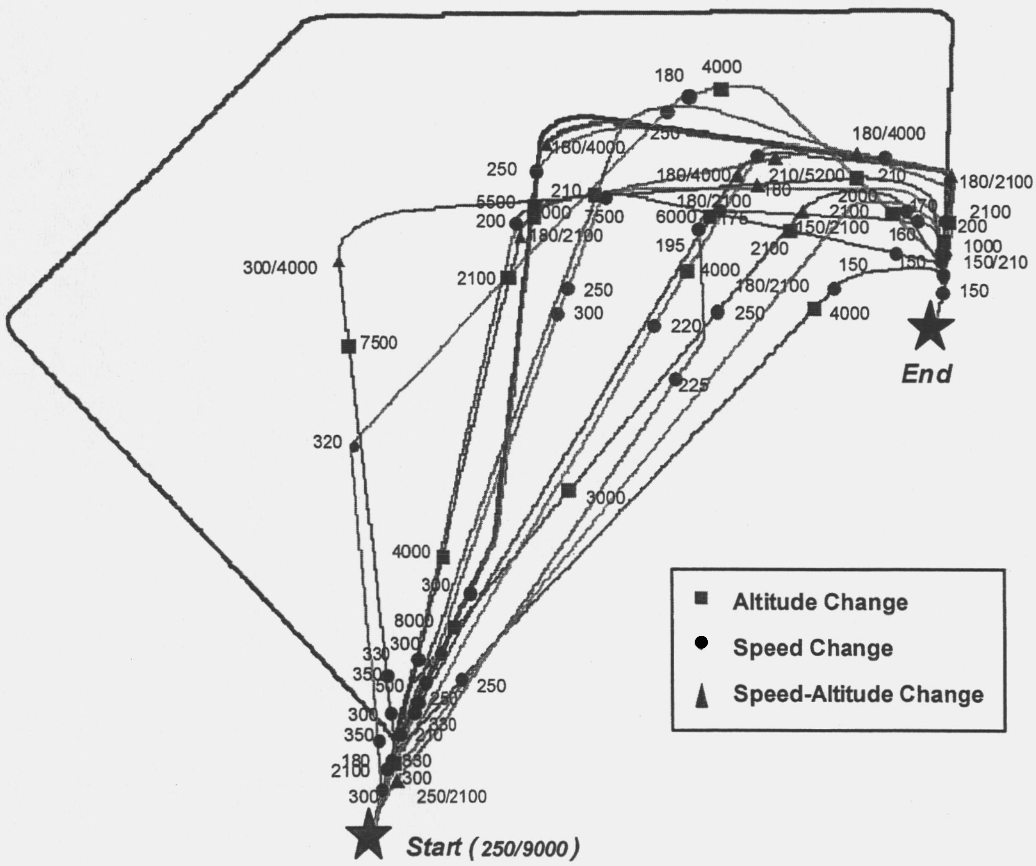


Figure 23 – Flight Paths and Altitude and Speed Changes in the Hydraulic Pressure Loss Scenario

Twelve of 16 pilots were at a point abeam the marker at the landing speed and glideslope altitude from where they started their turn onto final approach. The remaining four gave themselves a little more time by taking a turn further out from the marker and descending during the turn. This, however, resulted in three of the pilots reaching the outer marker (termination point) at an altitude higher than glideslope intercept altitude. The fourth pilot, due to high speed at the turn, did not have sufficient time and distance to slow down to the outer marker speed constraint. He did make the altitude constraint but could not line up for approach and was a little offset from the course. It was also observed that speed and altitude changes were made in no particular order, except that one was made only after the other was established. It was also seen that, in almost all the cases where a turn onto the base leg was required, pilots maintained a high speed up to a point abeam the marker and had shallow turns onto final approach.

An analysis of the subjective questionnaires revealed that the most common factors considered during re-planning were the distance to go, weather, time, and aircraft safety (Figure 24). In general, all pilots said that the first priority was to minimize the time of flight and the distance to go, irrespective of the situation. Other considerations included factors such as aircraft performance, safety of the maneuver and efficiency (Figure 25).

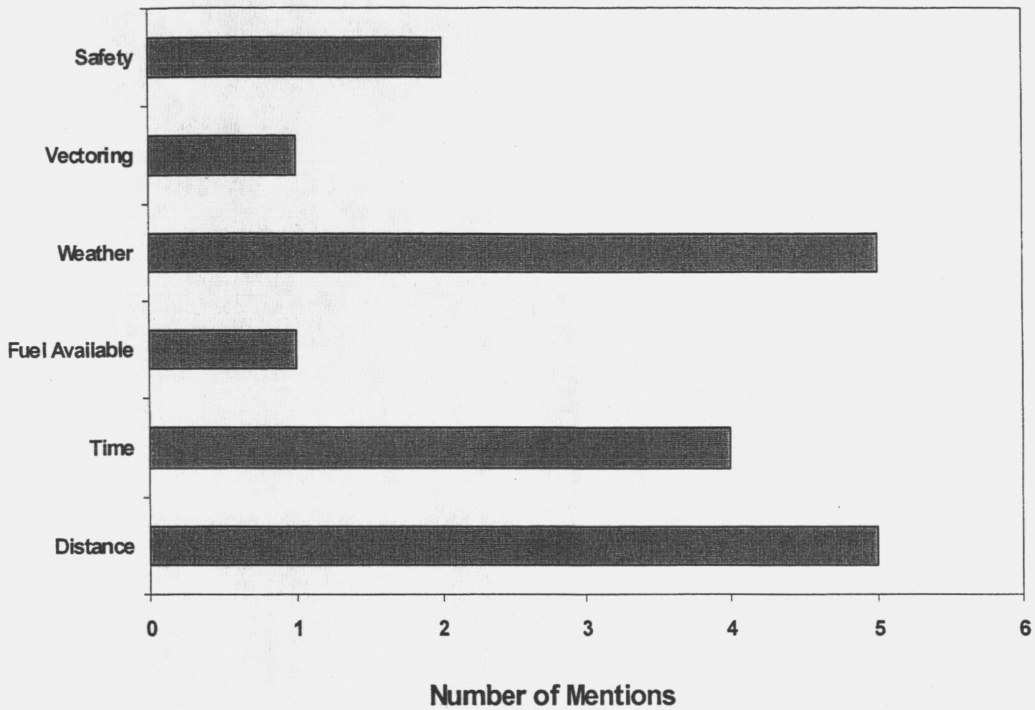


Figure 24 - Factors Considered During Planning

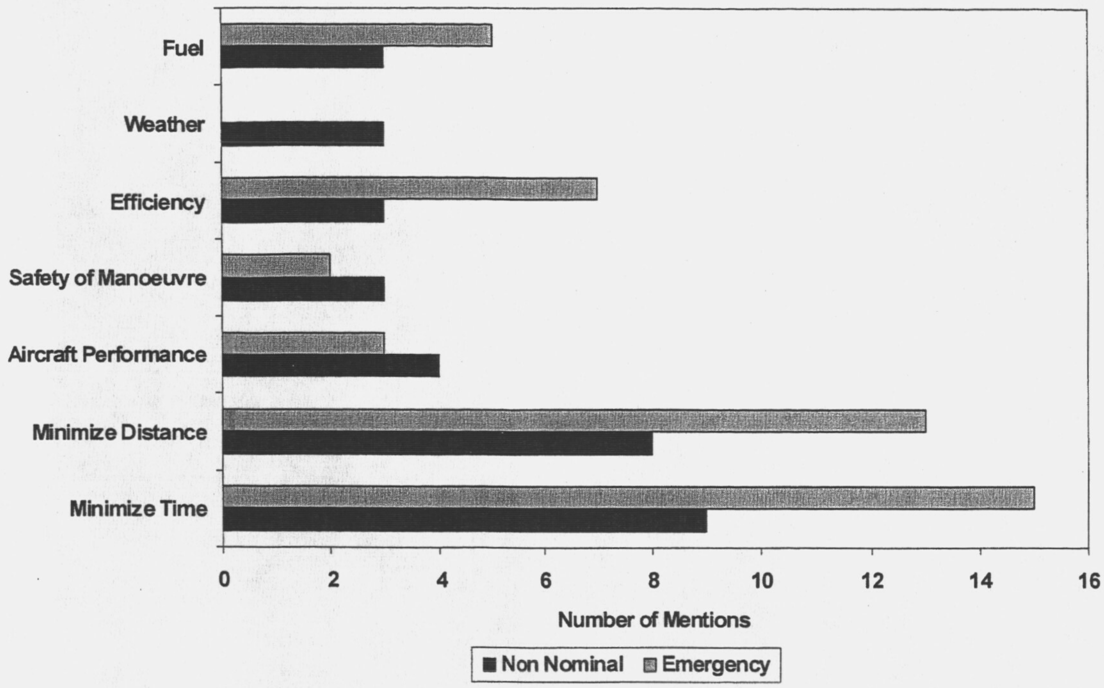


Figure 25 - Strategies in Choosing the Route Planned and Implemented

Pilot Planning Using the Mode Control Panel

When comparing the flight paths for the scenarios types (Appendix D.1), it was seen that, when using the MCP for the weather disturbance, two pilots went right of the weather and two pilots went left of the weather. Pilots who went right of the weather said it was easier to line up for approach and did not require any adverse maneuvering. In the restricted airspace scenario, it was seen that three pilots went right of the restricted areas and one went left. In the remaining non nominal scenarios, all the pilots followed similar right downwind paths. For the emergency scenarios, all pilots took the same right downwind and base leg paths. However, average speeds for the emergency scenarios were distinctly higher than for the non nominal scenarios, with two pilots flying a substantial length of the run at 400 knots in the medical emergency scenario.

Heading select (HDG SEL) was the most frequently used mode. HDG SEL was engaged from 1 to 5 times per run. Related to the usage of HDG SEL was the usage of the heading hold (HDG HLD) mode. This measure showed that pilots engaged this mode once on average for emergency (ranging from 0 to 4) and non nominal scenarios (also ranging from 0 to 4). Though the scenario did not show any significant effect on the usage of this mode, the average mode engagement for these two modes was higher for the emergency scenarios. The only significant factor here was the pilot, which suggests that the use of these modes for lateral navigation is more a personal choice.

A comparison of flight level change (FLCH) and vertical speed (V/S) modes showed that V/S mode proved to be a preferred mode for vertical navigation. It was observed that 2 of 16 pilots did not engage the FLCH mode in either of the scenarios in which they use the MCP. The specific scenario also did not affect their choice as was revealed in discussions during the experiment. The reason given was that they like to have control over the descent rates which can be defined in the V/S mode, but is internally calculated by the autopilot in the FLCH mode. None of the main effects had any significant effect on these two modes, which suggests that usage of these modes is a personal choice of pilots. Six pilots said that for emergencies, they preferred to use the more aggressive FLCH for climb and descent maneuvers. Seven pilots said they preferred V/S as it allowed them to control their own rate of descent/climb, though it did increase workload and monitoring activities slightly. The remaining three pilots did not give any preference in using these modes for vertical navigation.

Pilot Planning Using the Control Display Unit (CDU)

In the CDU (and its variants) cases, each click of the EXEC button was logged. The EXEC button was required to be pressed every time a change to the route was to be entered as the active route to be followed by the FMS. Specifically, these changes included adding/deleting a waypoint, erasing the previous action, resolving a route discontinuity, and making a speed or altitude modification. This was useful in analyzing the number of times that the plan was updated, and how thoroughly the pilots planned their task, i.e., whether they formed a skeletal plan and refined it along the way or took a little more time and proceed to implement a more concrete plan with fewer modifications.

When the timing of the EXEC button hits was looked at, it was seen that pilots who took longer to start and execute their plans had a spate of modifications and executions in the initial part of their plan and consequently fewer modifications along the way. This did reinforce the inference that the pilots who took longer to start planning had a more concrete idea of their planned route than other pilots. In addition to the above, it was observed that 10 of 16 pilots updated their plans more frequently in the non nominal scenarios than the emergency, five pilots updated their emergency plans more frequently and one showed no difference.

Pilot Planning Using the CDU with Autoplan Available (CDU+)

From a comparison of the flight paths for both scenario types, it was seen that, for the weather disturbance, only one pilot intentionally decided against the Autoplan and went to the right of the weather. The reasons given were the ease of lining up for approach and that it was a non nominal scenario. In the runway change scenario only one pilot followed the Autoplan route (with a few modifications) on its right downwind path, simply agreeing with the route in general and assuming that Autoplan gave the best route.

Pilots' reliance on the Autoplan was examined by the number of runs in which the Autoplan was the active route at the point when the run was terminated. When using the CDU+, one pilot did not use the Autoplan at all. Additionally, 7 of 16 pilots were seen to have used the Autoplan for both scenarios. From the remaining eight pilots, four used the Autoplan only for the emergency scenarios and four others used it only for the non nominal scenarios. In all runs where the Autoplan was used, modifications were made for a more direct route and to the speeds and altitudes in no particular order.

Pilot Planning Using the CDU with Autoplan Active at Start (CDU++)

From a comparison of the flight paths for the non nominal and emergency scenario types, it was observed that, for the non nominal scenarios, only one pilot (runway change scenario) chose to follow a route different to the others. In this case, however, the pilot simply followed the Autoplan route assuming it was the best route. From discussions and responses, it emerged that only distance and time were the important factors taken into account. In the remaining three non nominal scenarios, all the pilots followed similar downwind patterns with the corresponding turns to base leg. In the emergency scenarios, only one pilot followed a different route (fuel filter scenario).

In all runs using CDU++, the Autoplan was maintained as the active route up to the end of the run. However, changes were made to get a more direct routing, and also to speeds and altitudes to ensure a safe and expeditious flight. Another observation made here was that 9 of 16 pilots made substantial modifications to the Autoplan to the extent that they followed a different downwind path compared to the Autoplan.

Pilot Interaction with Automation

Pilot responses and simulator log files also gave us insight into the reliance of the pilots on the Autoplan. In general, all pilots with the exception of two agreed with the general routing that the Autoplan provided, but also concurred that they required speed and altitude changes and, in some emergency cases, quite extensive changes. However, they did approve of the Autoplan feature. The sole pilot who did not like the Autoplan feature did categorically state that he was not a big fan of automation as he did not agree with the extent to which it delegates control of the aircraft away from him. Figure 26 shows the number of runs each pilot had with the Autoplan as the active route.

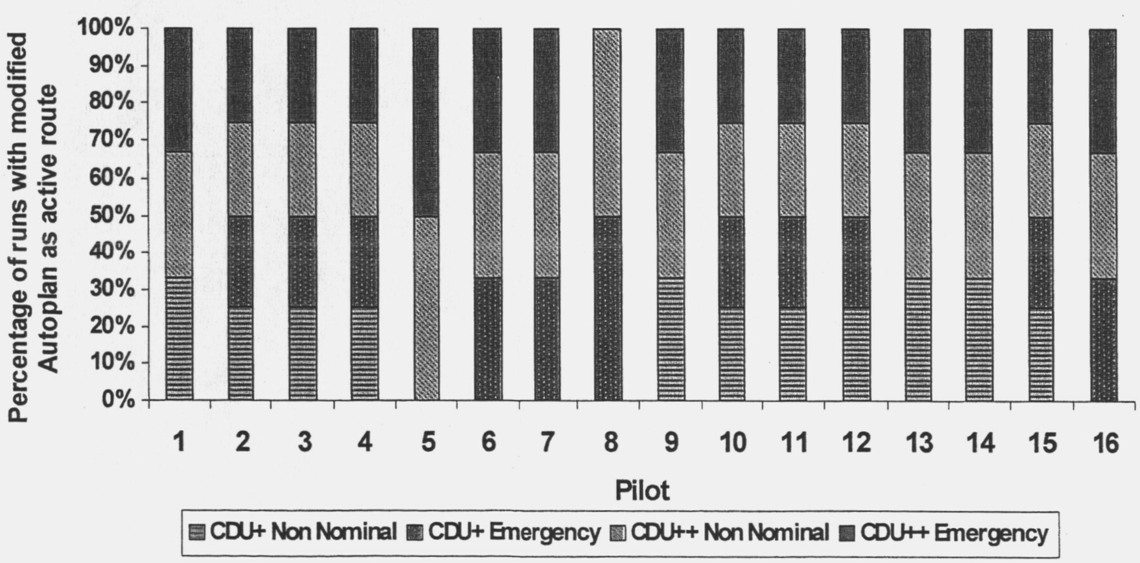


Figure 26 - Number of Runs with Modified Autoplan Active until End of Run

On the completion of each scenario, pilots were asked a series of questions pertaining to replanning in that scenario using the type of automation they used. Among the questions asked was a comparative evaluation of the automation used to what they would have preferred to use for that scenario on a Likert scale from 'Easier' to 'More Difficult' (Figure 27). It should be noted that, in each type of automation, there are a total of 32 runs with 16 pilots undergoing 2 runs each, one for each scenario type. These include cases where in a pilot may have preferred a different type of automation for each scenario type.

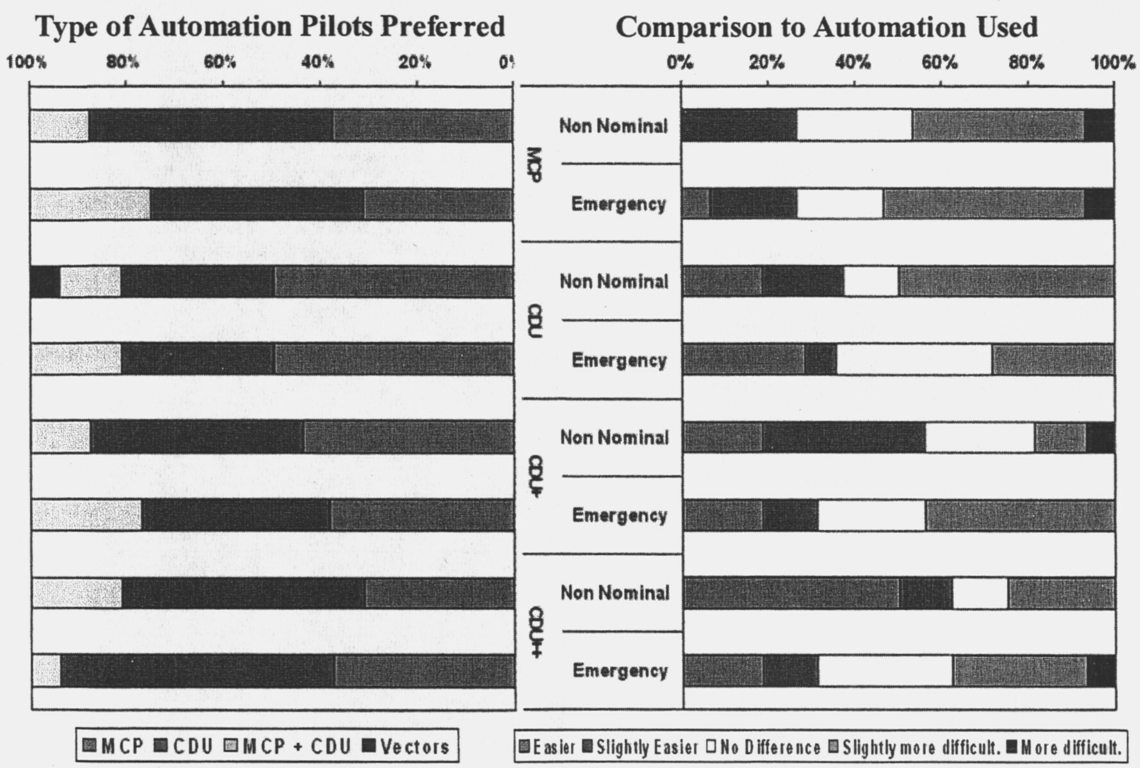


Figure 27 - Pilot Comparison of Automation Used with Preferred Automation

An interesting read from the above figure is that some pilots preferred to use a particular type of automation even though it resulted in more work for them and planning was more difficult. This could arise out of familiarity with the system currently being used and how often pilots use these in real world situations.

At the end of the experiment, the pilots were asked to rank the planning tools available to them; from best (1) to worst (4), according to which one was they felt was more useful for each scenario type. Figure 28 summarizes the rankings.

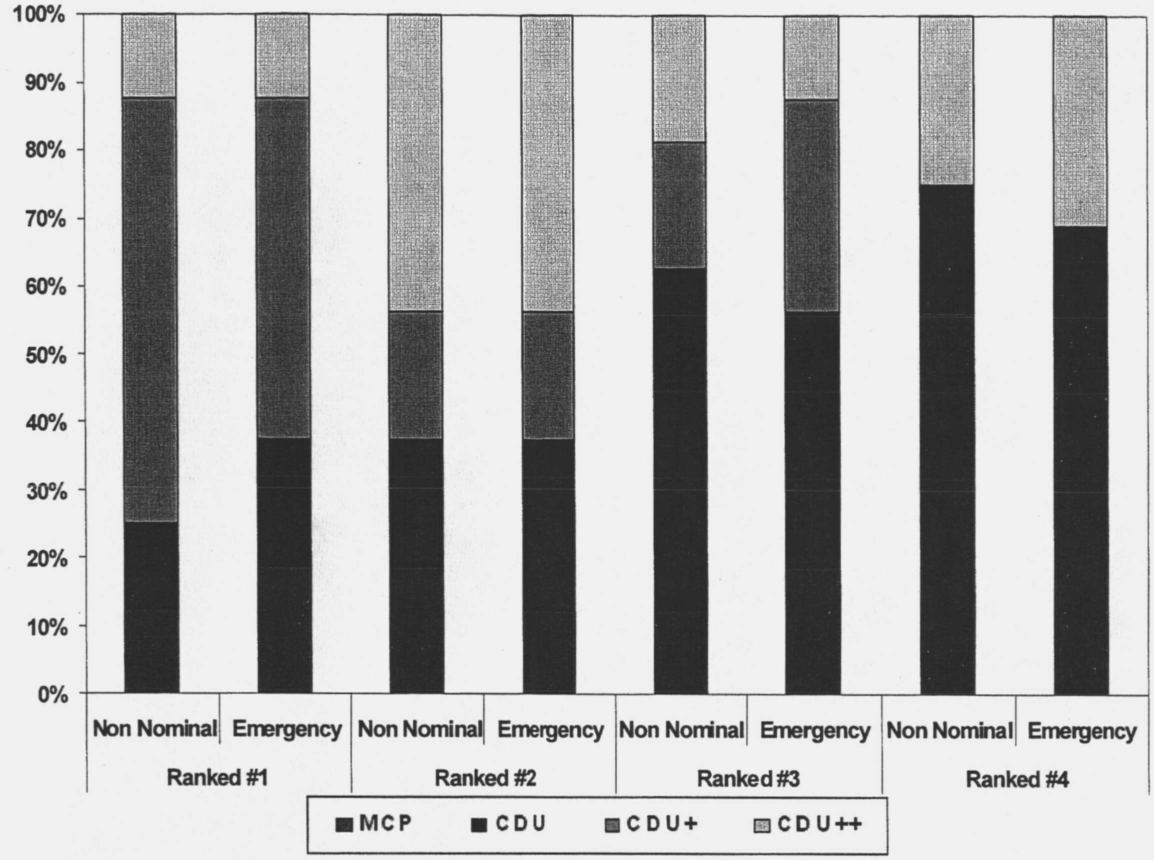


Figure 28 - Pilot Rankings of Automation Types per Scenario Type

From Figure 28 it was quite apparent that the CDU+ was the automation preferred by the pilots in the experiment with 62.5% of pilots rating it the best for the non nominal scenarios and 50% rating it the best in the emergency scenarios. Interestingly, 56% of pilots rated the MCP the worst for the non nominal scenarios and 50% rating it the worst in emergencies. A Wilcoxon signed ranking test (non parametric) was performed on the above response for both scenario types. In both scenario types, the CDU+ was rated as the best type of automation.

Finally pilots were also asked to describe the performance of the Autoplan. This was not specific to any scenario type. Table 6 shows the response of each pilot to the question: *"How would you describe the performance of the Autoplan? Please elaborate."* , and it can seen that all but two pilots approved of the Autoplan function although some also stipulated caveats such as wanting to double check the route it suggests.

Table 6 - Pilot Responses to Performance of Autoplan

How would you describe the performance of the Autoplan? Please elaborate.	
Pilot 1	It gave a very viable option that you could choose or reject. It would save effort and thought process if it was elected
Pilot 2	I found Autoplan very easy to use and it made my workload much less.
Pilot 3	Autoplan is a great idea if implemented correctly. It needs the ability to pick waypoints that are likely to be used in a given airspace. I think this could be accomplished in part by surveying ATC and having them suggest alternate route in their airspace. Another constraint is CDU memory, which is in short supply in the 757/767s I fly. As long as Autoplan has the ability to pick a logical, likely route, it will be a good thing. If however, it picks routes that will not be used in real life, it will become a button that never gets used.
Pilot 4	Helpful as a suggestion, that can be easily modified. Adds fixes that can be used without typing.
Pilot 5	Autoplan has no way of knowing what the objective is. Therefore, it may offer a long route when a short route is desired. I believe in most cases, I would not use Autoplan.
Pilot 6	I would not have picked most of the routes it did. A little aggressive for passenger operations and routes were longer.
Pilot 7	I liked Autoplan. Not sure that I wanted it to switch to it automatically (CDU++), but I found the displayed alternate route very helpful in picking the route I would use.
Pilot 8	Coupled with the visual representation, it provides me with great options; however, I am concerned about ATC's ability to go along with the plan.
Pilot 9	It may offer a good solution, then again it may not. Autoplan is not the best solution in all cases but at least look at it to evaluate it
Pilot 10	Good. It gave a quick route with an appropriate lead into final.
Pilot 11	Good. It gives a viable routing to destination and allows you to refine as necessary.
Pilot 12	I think it can be a useful system because it can save cockpit workload. It depends on how closely it would match optimum route and how likely pilot could stay on that route and not be altered by ATC.
Pilot 13	Generally good, but needs to be modified based on current factors.
Pilot 14	I think the Autoplan is a great tool, but it needs to be treated only as a tool to help me make rerouting decisions.
Pilot 15	It provided a shorter route to the airport. However ATC usually does the same to the extent that traffic allows.
Pilot 16	In general good. In time critical situations, it can give a good plan quickly and then you can take time refining it.

Workload Assessment

To assess the workload involved in each scenario, the pilot was asked to complete NASA Task Load Index (TLX) ratings at the end of every run. The worksheet probed the pilot for their personal assessment of workload on a continuous scale. Workload itself was broken down into 6 categories: mental demand, physical demand, temporal demand, performance, effort, and frustration. Workload within each type of automation is shown in Figure . For each of the above categories, a general linear model was fitted to examine the main effects. Subsequent ANOVA test were done where applicable. Categories which failed normality conditions were subjected to non parametric tests to examine any differences within the independent variables.

In the mental demand category, the residuals of the general linear model failed the Kolmogorov-Smirnov normality test ($p>0.150$), thus disallowing an ANOVA. Subsequently, a non parametric test, the Kruskal-Wallis Test was performed which showed no significant effect of scenario ($H = 1.76$, $P = 0.972$) or type of automation ($H = 0.94$, $P = 0.815$) on mental demand.

As with mental demand, physical demand also failed the Kolmogorov-Smirnov normality test ($p>0.15$), thus rendering the ANOVA test unusable. Similarly, a non parametric test revealed no effect of automation or scenario on physical demand, but run order effects were found. However, discussions with pilots and observations during the experiment revealed that any physical demand was more a result of using a virtual graphical user interface than of planning. Temporal demand also failed the normality conditions, when fitted to a general linear model. A non-parametric test revealed no significance of the main effects on this measure. The performance self-rating showed a significant automation effect ($F=15.81$, $p<0.001$). A 95% confidence Tukey test was further performed, which revealed that the MCP had the worst effect on performance. Pilot ratings of their effort were not affected by the automation types. However, the specific scenario did show a significant effect ($F=4.33$, $p=0.040$) on effort. Specifically, the weather disturbance scenario and the restricted airspace scenario had the most effect on effort among all the scenarios. Frustration was generally low and none of the variables showed any significant effect on the frustration level experienced by pilots during the task. Pilots may have reported a low frustration level in using the different autoflight systems since they have been exposed to these systems in real aircraft. Subsequent non parametric tests also failed any appreciable difference in any of the main effects.

The average workload rating (unweighted average of the TLX sub-scales) did not vary much with scenario type. In fact these were more specific to the type of automation wherein the workload rating for both scenario types was highest for the MCP and the lowest for the CDU+ (Figure 30) but was not statistically significant.

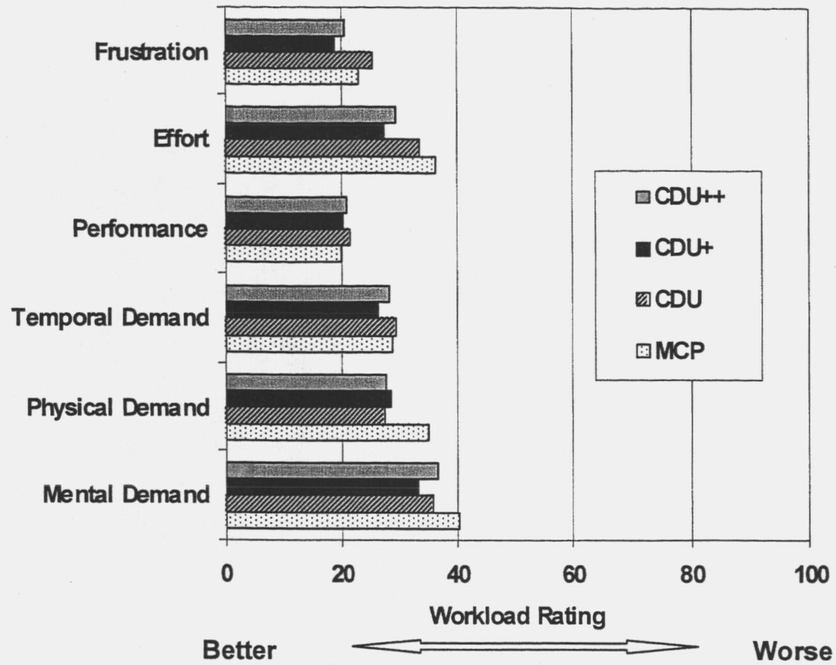


Figure 29 – Average TLX Workload Ratings for the Planning Task

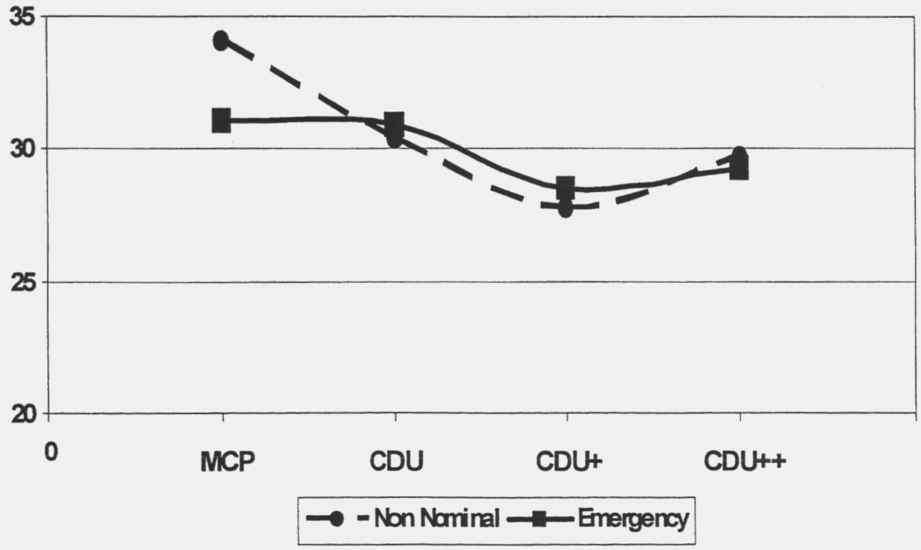


Figure 30 - Average Total Workload

'Faulty Autoplan' Scenario

The faulty Autoplan scenario that was run after completion of the first eight runs provided insight into the effect that a faulty Autoplan may have on the pilot's performance. Specifically the CDU++ generated a faulty plan which the aircraft would immediately start to follow at the beginning of the scenario. The Autoplan was erroneous in that, in a non nominal scenario type, it would provide a plan that was extremely aggressive and not safe for normal airline operations, i.e., it would generate an over aggressive plan that was suitable for a critical emergency. Likewise, for an emergency scenario type, it would generate a more circuitous route unsuitable for emergency, thereby ignoring the primary measures of time and distance. Eight pilots ran the ninth run in the non nominal scenario and eight pilots ran it for the emergency scenario, thereby giving 16 runs (data points). Figures 31 and 32 show the flight paths of the pilots in this run for the two scenario types.

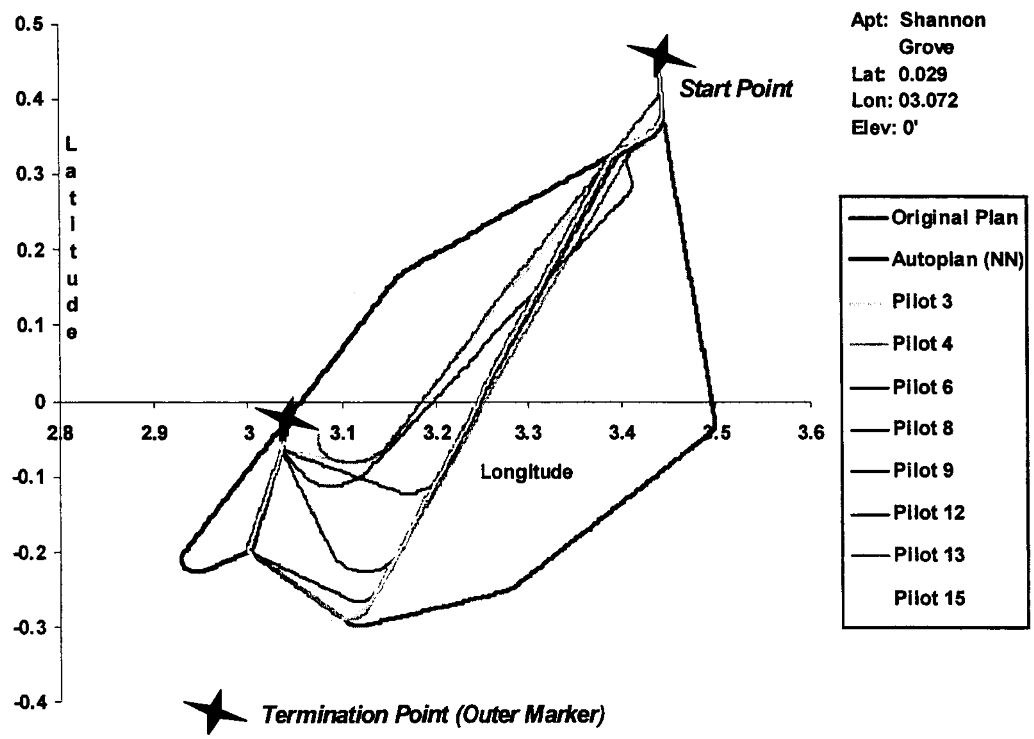


Figure 31 - Flight Paths for Non Nominal Faulty Autoplan Scenario

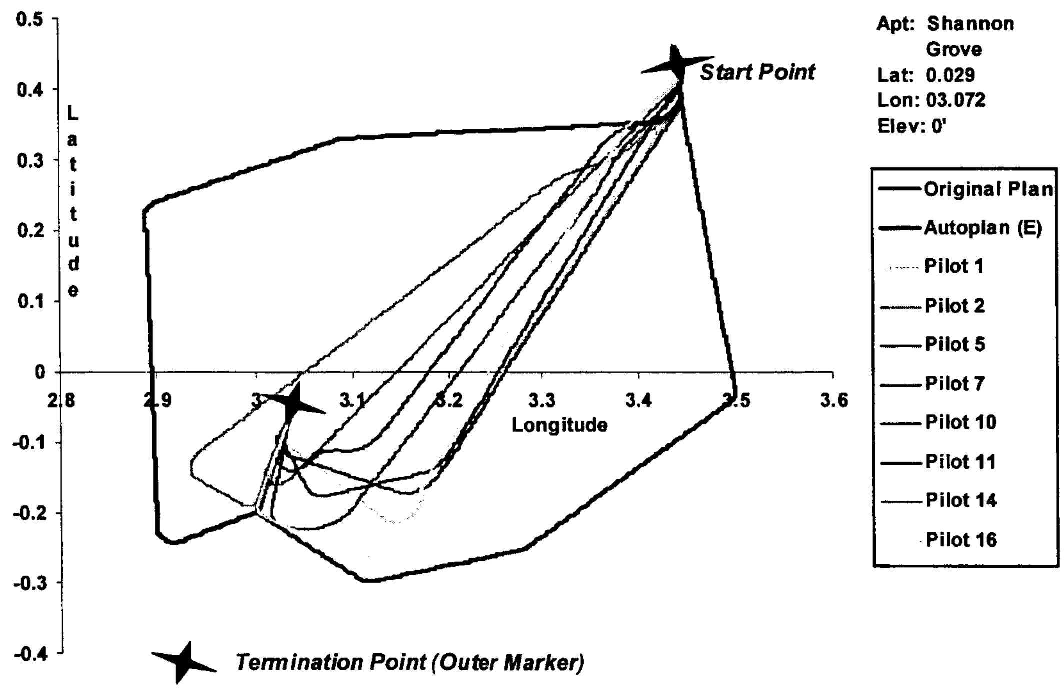


Figure 32 - Flight Paths for Emergency Faulty Autoplan Scenario

Although there are an insufficient number of data points for a statistical analysis, some trends merit discussion. Regardless of scenario type, pilots' primary aim was to minimize time aloft and distance to travel. In the non nominal scenarios six out of eight pilots did not activate the original route (RTE 1) but chose to modify the Autoplan. The two pilots that did activate the original route took, from the start of the run, an average time of 2.131 minutes to activate and 2.489 minutes to start modifying the route (the other six pilots took an average time of 1.261 minutes to start modifying the route). This suggests that they did evaluate the two plans available.

It was observed that in cases where pilots activated the original route, the average number of modifications to the active plan was six whereas, for the other six pilots, the number of modifications increased to eight, thereby suggesting that the original route was better than the Autoplan and required less modification, which was subsequently verified through observations and comments made by the pilots during the experiment and debriefing.

The number of modifications to the plan was measured by the number of times the pilots pressed the EXEC button to execute a modification. In the emergency scenarios five out of eight pilots did not activate the original route (RTE 1) but chose to modify the Autoplan. The three pilots that did activate the original route, however, took an average time of 0.483 minutes to do so. This suggests that they immediately recognized the erroneous Autoplan and proceeded to activate the original plan. In the non nominal scenario, the average deviation in distance flown was 10.562 miles with a corresponding saving in flight time of 7.027 minutes. For the emergencies these measures corresponded to 10.295 minutes saved in flight time and 25.819 miles saved in flying distance. These were measured by the difference in times of the modified route and the unmodified active route. Figures 33 and 34 show a snapshot of the various measures for the faulty Autoplan scenario:

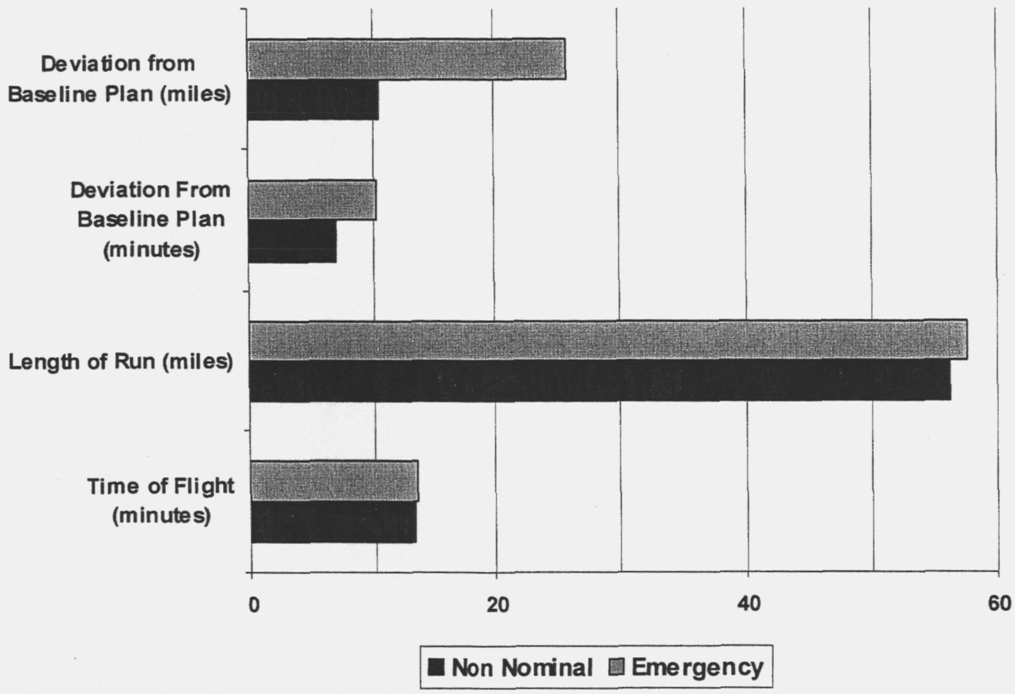


Figure 33 - Planning Performance Measures in Faulty Autoplan Scenario

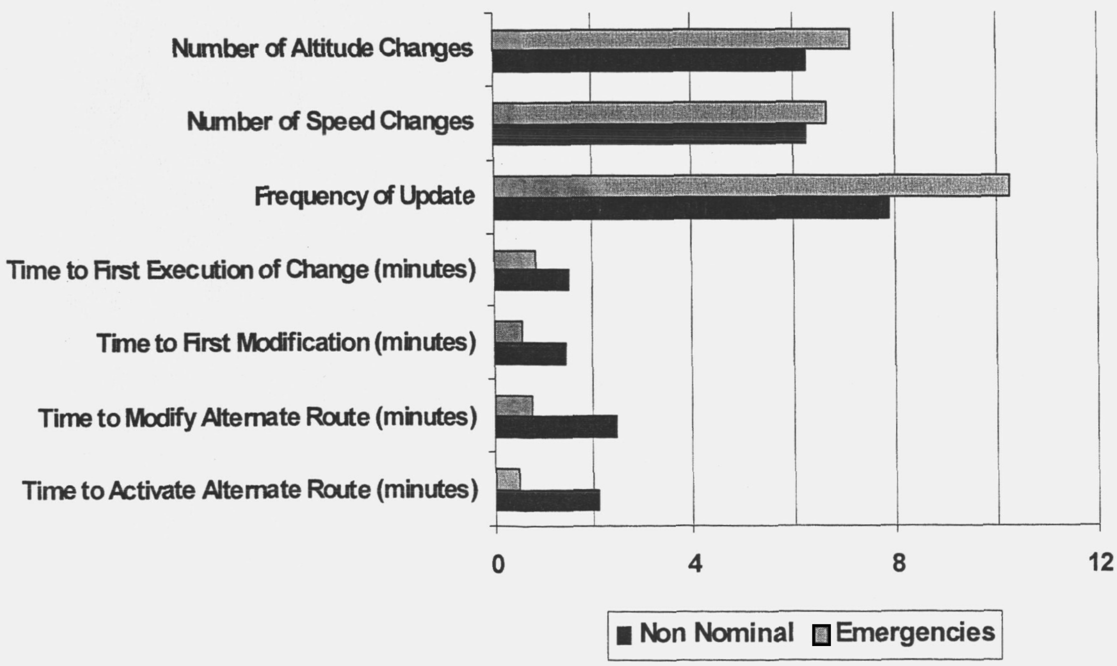


Figure 34 – Planning Behavior Measures for Faulty Autoplan Scenario

Discussion and Conclusions

The results of this experiment suggest that pilot behavior and performance differs in different situations, be it non nominal or emergency. When pilots used only the MCP, the time of flight and the length was lower (i.e., better) than with the other types of automation. With the MCP, the emergency scenarios showed markedly lower values for the primary measures of performance than the non nominal scenarios, which had a stronger effect on the safety of the flight. This was attributed mainly to the fact that pilots did not need to spend too much time creating and modifying fixes, but rather spent more time on speed and altitude changes.

The CDU only, however, showed a slight degradation in pilot performance. The workload assessment showed no significant difference from that with the MCP, but the primary measures of time of flight and length of run were the highest in this type of automation. Average deviations were about the same as that of the MCP suggesting that resulting plans were similar, but the comparative flights varied substantially. The non nominal scenarios show higher averages for time and distance than the emergency scenarios; however, these also show a markedly higher average for the emergency scenarios when compared with the MCP only case with no appreciable change in overall workload. This case did show a higher level of frustration than the other automation types mainly because pilots had to spend time entering and modifying fixes, and, in some cases, pilots had been previously been exposed to the other variants of the CDU.

The variants of the CDU, namely CDU+ and CDU++, were well received by the pilots because of the additional Autoplan feature which was found to tremendously reduce pilot workload during replanning. Though the CDU+ did show a relatively higher temporal demand for both scenario types, it showed overall a much better performance in reducing time and distance and the subsequent total workload.

It was also seen that with all the variants of the CDU, pilots made substantial changes to the Autoplan. The Autoplans were created to meet mind airspace regulations; however, the inability of the plan to take advantage of the air traffic controller giving the pilot discretion over the route explained the changes made by the pilots to the Autoplan. These factors highlight the need for careful design of the Autoplan generator to be context sensitive including the ability to generate plans for both non-nominal and emergency situations and to take advantage of relaxed ATC restrictions. Pilot comments concerning the performance and usefulness of the Autoplan were more favorable than indicated by the performance measures. Indeed, most of the pilots believed that the Autoplan could be useful but at the same time expressed a number of concerns about its implementation.

The results of this preliminary experiment suggest that the functional concept of an automatically generated plan is an endeavor worth pursuing which provides the pilot much needed assistance in replanning a flight route. Additionally, it was observed that pilots tended to think of plans as a two dimensional space at any time.

Although this research specifically studied airline pilots' planning behavior in glass cockpit using current autoflight interfaces (MCP and CDU), the results suggest several broader implications for cockpit planning aids in general. The most important is the level of intelligence required by the FMS to generate such a plan on its own. While most pilots did say it was useful, some shot down the idea on the ground that they preferred to either create their own plans (even if it increased workload a little and increased time), or hand fly the aircraft as it afforded more control of the aircraft. In this experiment, for example, some pilots pointed out that the Autoplan did give a very good initial routing with minimal route changes, but was not very effective with the speed and altitude management.

Successful implementation of such a concept is highly dependent on the level of artificial intelligence, context sensitivity to and the sensing of external factors such as traffic, weather and terrain. The objective function or the goal of the plan should coincide with the specific situation at hand, be it non nominal or emergency in nature and whether the aircraft has been compromised or not.

Some pilots also observed that such a concept would be more useful in an en-route environment as terminal area traffic control is far stricter and more stringently regulated. A more dynamic and real time update of the plan would also be useful with additional information in the form of ETA to active waypoints would also be helpful to pilots.

Perhaps a more important question is the location of such a system. Though the Autoplan (and subsequently the CDU+) did not have much effect on pilot behavior, it could be located in the aircraft FMS or used by air traffic control level to create better aircraft routings, perhaps updated when the situation changes.

Other additions that may prove to be helpful are a complementary display which shows a vertical and horizontal display of the Autoplan in a space relative to other routes and traffic, weather, and terrain, as well as supplemental information of estimated time to travel, estimated distance to go, estimated fuel consumption and savings on time and distance compared to the previous plan. These additions, with subsequent testing, can better confirm the effectiveness of such a concept.

With the development of free flight, the concept of an automatic plan generator would greatly enhance in-flight re-planning tasks and could have better context sensitivity if, in addition to the above mentioned enhancements, Autoplan could incorporate 'Party Line' Information (PLI) such as real time and current pilot reports about weather and traffic, Collaborative Decision Making (CDM) information such as airspace system status, equipment availability and weather, and the output of other tools such as the User Request Evaluation Tool (URET) for conflict prediction, passive Final Approach Spacing Tool (pFAST) for terminal area arrival and departure streaming operations, and Traffic Management Advisor (TMA) for en-route traffic management.

Results also bring into focus the effectiveness of the control display unit as a replanning interface. With its text display and keystroke method of data entry, planning interfaces such as touch screens which allow pilots to graphically pick waypoints and define a flight path may prove to be both easier to use and facilitate pilots in creating better plans. Such a system does not call for elimination of the CDU from the flight management system, but current methods of using planning interfaces in flight decks do call for a more efficient interface. Such a system would be efficient in that pilots would have the system in front of them (thus allowing pilots to monitor other flight instruments simultaneously), reduce physical movements in terms of data entry into the FMS and not requiring the pilot to constantly go heads down when creating the plan and then looking up to verify the plan.

Summary

Prior research has, even in relatively low-fidelity and simple simulator tests, found situations where pilots had difficulty in planning and flying a safety trajectory in an emergency. Likewise, early results suggest that intelligent cockpit system can potentially aid pilots at this difficult task, if they are imbued with the correct knowledge and logic. This project has outlined a series of multi-disciplinary research tasks which lay the foundation for these knowledge and logic requirements, and test their efficacy through the development and simulator testing of a prototype cockpit system, and through examination of other applications of these research gains in procedure development.

While these results help specify a flight capable system to aid pilots in emergency situations, they also may have a far-wider impact than the isolated development of one system. The knowledge gained provide researchers and designers with a better theoretical understanding, based on empirical research, of how pilots plan trajectories, use procedures, and react to emergencies. From this, insights have been made on the representation of procedures, and on the use of automated systems in aeronautics.

These results are also being prepared for broader dissemination in the scientific community by publishing in peer-reviewed technical journals.

References

Chen, T. L., & Pritchett, A. R. (2000). Testing of an on-board emergency flight planner. Proceedings of the Digital Avionics Systems Conference.

Chen, T. L., & Pritchett, A. R. (2001). Cockpit decision-aids for emergency flight planning. Journal of Aircraft.

Chou, C.-D., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. The International Journal of Aviation Psychology, 6*(4), 307-320.*

deBrito, G. (1999). Human Error in Procedure Following: A First Classification. Proceedings of the Human Error, Safety and Systems Development, *Presented.*

Dien, Y., Montmayeul, R., & Beltranda, G. (1991). Allowing for human factors in computerized procedure design. Proceedings of the Human Factors Society 35th Annual Meeting, *639-642.*

Fan, T. P., Hyams, D. S., & Kuchar, J. K. (1998). Study of in-flight replanning decision aids. Proceedings of the AIAA Guidance, Navigation, and Control Conference, *980-988.*

Funk, K. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. The International Journal of Aviation Psychology, 1*(4), 271-285.*

Guerlain, S. A., Smith, P. J., Obradovich, J. H., Rudmann, S., Strohm, P., Smith, J. W., Svirbely, J., & Sachs, L. (1999). Interactive critiquing as a form of decision support: An empirical evaluation. Human Factors, 41*(1), 72-89.*

Hutchins, E. (1995). Cognition in the Wild. *Cambridge, MA: The MIT Press.*

Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. Human Factors, 36*(1), 94-119.*

McCarthy, J. C., Wright, P. C., Monk, A. F., & Watts, L. A. (1998). Concerns at work: Designing useful procedures. Human-Computer Interaction, 13*(4), 433-457.*

Mosier, K. L., & Skitka, L. J. (1999). Automation use and automation bias. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, *344-348.*

Ockerman, J. J., & Pritchett, A. R. (2000). A Review and reappraisal of task guidance: Aiding workers in procedure following. International Journal of Cognitive Ergonomics, 4*(3), 191-212.*

Ockerman, J. J., & Pritchett, A. R. (2004). Improving performance on procedural tasks through presentation of locational procedure context: An empirical evaluation. Behaviour and Information Technology, 23*(1), 11-20.*

Pritchett, A. R., Nix, D. C., & Ockerman, J. J. (2001). Empirical evaluations of pilot planning behavior in emergency situations. Proceedings of the 4th Annual Meeting of the Human Factors and Ergonomics Society.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. Cognition and Instruction, 1*(4), 359-396.*

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. IEEE Transactions on Systems, Man, and Cybernetics--Part A: Systems and Humans, 27*(3), 360-371.*

Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. *Cambridge: Cambridge University Press.*

Swezey, R. W. (1987). Design of job aids and procedure writing. In G. Salvendy (Ed.), Handbook of Human Factors *(pp. 1039-1057). New York, NY: John Wiley & Sons.*

Vicente, K. J. (1999). Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. *Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.*

Wright, P., Pocock, S., & Fields, B. (1998). The prescription and practice of work on the flight deck. Proceedings of the Ninth European Conference on Cognitive Ergonomics, *37-42.*